Canadian Mathematical

BULLETIN

Canadien de Mathématiques

Vol. 25 No. 2 1982 JUNE

Jeffrey–Williams Lecture, 1981

A group theoretic approach to the equations of Plasma Physics .... J. E. Marsden 129

A note on completely and absolutely monotone functions .... A. Mahajan and D. K. Ross 143
Five mutually tangent spheres and various associated configurations ..... J. F. Rigby 149
Integral representation by boundary vector measures .... P. Saab 164
Topologically indexed function spaces and adjoint functors .... S. B. Niefield 169
Girth and independence ratio .... G. Hopkins and W. Staton 179
Solvable groups of unipotent elements in a ring .... A. A. Klein 187
Notes on the Birkhoff algorithm for doubly stochastic matrices .... R. A. Brualdi 191
On the divisibility of the class numbers of $Q(\sqrt{-p})$ and $Q(\sqrt{(-2p)})$ by 16 .... P. A. Leonard and K. S. Williams 200
On Dunford–Pettis operators .... E. Saab 207
Uniformly Lipschitzian semigroups in Hilbert space ... D. J. Downing and W. O. Ray 210
A presentation for a group of integer matrices .... N. Smythe 215
Cyclotomic splitting fields .... R. A. Mollin 222

NOTES MATHEMATIQUES

Systems of Briot-Bouquet equations with analytic solutions .... D. Westreich and E. Podolak 230
A new proof of the Snake theorem .... L. L. Keener 234
On the number of equal partitions of $\{1, \dots, n\}$ into two sets of equal cardinalities and equal sums .... H. Prodinger 238
Discrete sets and discrete maps .... R. H. Warren 242
The extended center of coproducts .... W. S. Martindale, 3rd 245

PROBLEMES ET SOLUTIONS 249

ISSN 0008—4395

LA REDACTION

E. G. Goodaire et R. Piccinini	Rédacteurs-en-chef
E. J. Barbeau	Rédacteur-Problèmes
S. Swaminathan	Rédacteur-Gérant

LA POLITIQUE EDITORIALE

La politique éditoriale du Bulletin Canadien de Mathématiques est exigeante quant au niveau mathématique. Pour être publié dans le Bulletin un article doit être sans faute, original, bien écrit et doit intéresser une bonne section du public mathématique. La rédaction se réserve le droit de publier des articles, notamment s'ils sont courts, dans la partie *Notes*.

RENSEIGNEMENTS GENERALES

Il est éntendu que les articles soumis au Bulletin ne sont considérés que s'ils n'ont pas été déjà publiés ou bien soumis ailleurs. Le Bulletin publie surtout en français et en anglais. Envoyez deux copies aux rédacteurs-en-chef, Départment des mathématiques et statistiques, l'Université Mémoriale de Terre Neuve, St. Jean, Terre Neuve, Canada A1B 3X7. Addressez les soumissions destinées à la section de Problèmes à E. J. Barbeau, Départment des Mathématiques, l'Université de Toronto, Toronto, Ontario, Canada M5S 1A1. Le prix de l'abonnement annuel (4 numéros) du *Bulletin* est de $30; les membres des sociétés de mathématiques reconnues auront droit à un prix réduit de $15. Prière d'adresser les demandes d'abonnements à Graham P. Wright, Société Mathématique de Canada, 577 King Edward, Ottawa, Ontario, Canada, K1N 6N5. Tout autre correspondance à S. Swaminathan, Départment des mathématiques, statistiques et d'informatique, l'Université Dalhousie, Halifax, Nouvelle Ecosse, Canada B3H 4H8.

Pour couvrir partiellement les frais de publication, l'organisme representé par l'auteur sera prié de verser un montant de $30 per page imprimée. En aucun cas l'auteur ne sera personellement responsables de ces frais, ni l'acceptation d'un article subordonné à leur paiement. L'auteur recevra deux éxemplaires des épreuves, dont une à renvoyer, avec un bon de commande de tirages à part. On n'en envoie qu'à un auteur. Aucune modification ou correction autre que celles d'erreurs typographiques sur l'épreuve ne sera acceptée sauf accord spécial. Dans ce cas, elles feront l'objet d'une facturation à l'auteur.

En plus du *Bulletin*, on publie le *Journal Canadien de Mathématiques* qui parait tous les deux mois, dont le prix de l'abonnement annuel est de $48; les membres des sociétés de mathématiques reconnues auront droit à un prix réduit de $24. Priere d'adresser les demandes d'abonnements au gérant, University de Toronto, Toronto, Canada, M5S 1A1.

DROITS D'AUTEUR

Dès la réception d'un manuscrit, on enverra à chaque auteur un formulaire concernant les droits d'auteur. Son retour dûment rempli est nécessaire pour la publication du manuscrit.

La Société Mathématique du Canada donne la permission à un lecteur de cette publication de faire une copie d'un article à l'usage personnel. La copie en grand nombre n'est permis qu'avec la permission écrite du gérant.

© Société Mathematique du Canada, 1982

Le *Bulletin* reçoit CRSNG (Canada) une subvention aux coûts de publication.

Publié par la Société Mathématique du Canada
Imprimé à 'The Universities Press (Belfast) Ltd., Northern Ireland'.

For an English version of the above see the issue of
March or September.

Canad. Math. Bull. Vol. **25** (2), 1982

A GROUP THEORETIC APPROACH TO THE EQUATIONS OF PLASMA PHYSICS

JERROLD E. MARSDEN*

ABSTRACT. This paper concerns the interaction between group theory and classical mechanics in general and with the application of this theory to plasma physics in particular.

Mechanics has had a profound role in the historical development of mathematics. Leading original thinkers in pure mathematics such as Newton, Euler, Lagrange, Jacobi, Laplace, Cauchy, Gauss, Riemann, Poincaré, Hilbert, Birkhoff, Smale and Arnold were also great original thinkers in various facets of mechanics. The interaction between mechanics and pure mathematics remains one of the most active and flourishing areas of current research. It is this flavor which I hope to convey in this lecture.

Acknowledgements. Much of the work that I will describe below was done in collaboration with Alan Weinstein (see Marsden and Weinstein [1981]). A seminal paper of Morrison [1980] originally shown to us by Allan Kaufman was our starting point. Our ideas were developed in a series of lectures given by both of us to Kaufman's plasma physics-dynamics seminar early in 1981 at Berkeley. The participant's enthusiasm and encouragement were very important.

I thank the officers of the Canadian Mathematical Society for the opportunity and honor to present these ideas in the Jeffrey–Williams lecture.

I wish to dedicate this lecture to my undergraduate teachers from Toronto who had such a positive influence on my career, especially Professors Atkinson, Coxeter, Davis, Duff, Pillow, Rooney, Rund, Scherk, Scherk and Vanstone. It was through them that I first appreciated the fact that pure and applied mathematics have a magnificant intersection within mathematics.

§1. **Introduction.** Commencing around 1750, Euler discovered and investigated the basic equations governing two of the most important mechanical systems, namely the rigid body and a perfect incompressible fluid. He would have been pleased with the fundamental paper of Arnold [1966] which showed

---

Received by the editors September 20, 1981.

* This paper is based on the Jeffrey–Williams lecture given at the meeting of the Canadian Mathematical Society, Dalhousie University, Halifax, Nova Scotia, May 28, 1981 upon the invitation of the Research Committee of the CMS. Research partially supported by NSF grant MCS-78-06718, ARO contract DAAG-29-79-C-0086, and the Miller Institute for Basic Research.

AMS (1980) Classification: 58D, 58E, 58F, 70H, 70W.

This One
WT1W-J2A-PDAX

that both of these sets of equations are of a special Hamiltonian type which is associated with a Lie group; in fact the group may be regarded as the configuration space of the system. For the rigid body the group is $SO(3)$, the proper orthogonal transformations of Euclidean 3-space $\mathbb{R}^3$ and for a perfect incompressible fluid, the group is $\mathscr{D}_{\text{vol}}$, the group of volume preserving diffeomorphisms of a region in $\mathbb{R}^2$ or $\mathbb{R}^3$.

The paper of Ebin and Marsden [1970] showed how to put Arnold's ideas for a perfect fluid into a rigorous infinite dimensional context and showed that this procedure leads to new existence, and convergence theorems that are ultimately useful in numerical work (see Chorin et. al. [1978] and references therein). It was noted there that many of the results on $\mathscr{D}_{\text{vol}}$ extend to $\mathscr{S}$, the group of canonical transformations of a given symplectic manifold i.e. a given phase space.

Here we show that there is a fundamental system associated to $\mathscr{S}$ in the same way that the rigid body is associated with $SO(3)$ and fluids are associated with $\mathscr{D}_{\text{vol}}$. This system consists of the Poisson–Vlasov equations for a collisionless plasma and its generalization, the Maxwell–Vlasov equation. These equations are as fundamental to the plasma community as the Navier–Stokes equations are to the fluids community.

We list below some of the important mechanical systems that have as configuration spaces a Lie group; (some additional details are provided in the next section).

System	*Group*
1. free rigid body	$SO(3)$ = rotation group
2. heavy top	$E(3)$ = Euclidean group
3. perfect incompressible fluid	$\mathscr{D}_{\text{vol}}$ = volume preserving diffeomorphisms
4. compressible fluid	$\mathscr{D} \times \mathscr{F}$ = semi-direct product of diffeomorphisms and functions
5. Korteweg de Vries Equation	$\mathscr{I}\mathscr{F}$ = group of invertible Fourier Integral operators
6. Toda lattice	H = invertible lower triangular matrices
7. Liouville equation of any Hamiltonian system	$\mathscr{S}$ = canonical transformations
8. Heisenberg equation of quantum mechanics	$\mathscr{U}(\mathscr{H})$, unitary group of complex Hilbert space
9. Lax equations for non-linear waves	$\mathscr{U}$ = unitary group
10. Poisson–Vlasov equations	$\mathscr{S}$ = group of canonical transformations

There are other basic systems in mechanics whose connection with a Lie group is less direct. For these, there is a more general way to link Lie groups and mechanical systems via *symmetry and reduction* (Marsden and Weinstein [1974]). This procedure includes many additional interesting systems besides those already listed. A few are:

System	*Group*
11. Maxwell's equations	$\mathcal{G}$ = gauge group of electrodynamics
12. Yang–Mill's equations	$\mathcal{G}$ = automorphisms of a principle bundle
13. Einstein's equation of General Relativity	$\mathcal{D}$ = diffeomorphism group of space-time
14. Supergravity	S = group of supersymmetry transformations.

In the following section we shall explain the coadjoint orbit scheme which encompasses systems 1–10 and shall indicate how the Euler equations for a rigid body in system 1 and how systems 8 and 9 fit this scheme. Then we go on to explain how Maxwell's equations can be described in terms of reduction. Finally we describe system 10 and how it is coupled with Maxwell's equations to produce the Maxwell–Vlasov system.

It will be necessary to assume some background in modern mechanics and Lie groups. See for example Arnold [1978] or Abraham and Marsden [1978].

§2. **Coadjoint orbit structures.** Let G be a Lie group and $\mathfrak{G}$ its Lie algebra. Let $\mathfrak{G}^*$ be the dual space of $\mathfrak{G}$. Let $\mu \in \mathfrak{G}^*$ denote a typical point. Let $F:\mathfrak{G}^* \to \mathbb{R}$ be a smooth real-valued function defined on $\mathfrak{G}^*$. Define its *functional derivative* $\delta F/\delta\mu : \mathfrak{G}^* \to \mathfrak{G}$ by

$$DF(\mu)\cdot \nu = \left\langle \nu, \frac{\delta F}{\delta \mu}\right\rangle$$

where ν is an arbitrary element of $\mathfrak{G}^*$, $DF(\mu):\mathfrak{G}^* \to \mathbb{R}$ is the (Fréchet) derivative of f at μ, $\langle , \rangle$ is the pairing between $\mathfrak{G}^*$ and $\mathfrak{G}$ and $\delta F/\delta\mu$ is understood to be evaluated at μ.

Define a bracket on functions F, G by

$$\{\{F, G\}\}(\mu) = -\left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu}\right]\right\rangle$$

where $[,]$ denotes the Lie bracket of the Lie algebra $\mathfrak{G}$.

We call $\{\{ , \}\}$ the *KAKS bracket* after Kirillov, Arnold, Kostant and Souriau who used it in various forms in their work in representation theory, Hamiltonian systems on Lie groups and quantization.

The KAKS bracket is clearly antisymmetric, is bilinear in F and G and is a derivation in F and G. Somewhat less obvious, but true, is Jacobi's identity. Thus, the KAKS bracket makes $C^\infty(\mathfrak{G}^*, \mathbb{R})$, the C^∞ real valued functions on $\mathfrak{G}^*$

Copyrighted material

into a Lie algebra. One says that $\mathfrak{G}^*$ together with $\{\{,\}\}$ forms a *Poisson manifold*; c.f. Guillemin and Sternberg [1980].

If $H:\mathfrak{G}^* \to \mathbb{R}$ is a given (energy) function there is a unique vector field X_H on $\mathfrak{G}^*$ such that for any $F \in C^\infty(\mathfrak{G}^*, \mathbb{R})$,

$$DF(\mu)\cdot X_H(\mu) = \{\{F, H\}\}$$

The left side is $\dot{F} = dF/dt$, the rate of change of F under the vector field X_H. We call $\dot{F} = \{\{F, H\}\}$ the evolution equations determined by H in Poisson-bracket form, while $\dot{\mu} = X_H(\mu)$ are the KAKS evolution equations themselves. They are clearly equivalent descriptions.

The above development may seem pretty far removed from the classical Poisson bracket

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}\right)$$

and Hamilton equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}^i = -\frac{\partial H}{\partial q^i}.$$

In fact, through reduction we can derive the KAKS Poisson structure $\{\{,\}\}$ from the traditional Poisson brackets in a natural way. This can proceed via two equivalent routes:

ROUTE 1. Regard $\mathfrak{G}^*$ as T_e^*G, the dual of the tangent space to G at the identity $e \in G$. For $F, G:\mathfrak{G}^* \to \mathbb{R}$, extend F and G to functions $f, g: T^*G \to \mathbb{R}$ by left translation. Then $\{f, g\}$ is defined by the classical formula where (q^i, p_j) are cotangent bundle coordinates. Let $\{\{F, G\}\}$ be $\{f, g\}$ restricted to $\mathfrak{G}^*$. Then this yields the KAKS bracket.

ROUTE 2. G acts on $\mathfrak{G}^*$ by the co-adjoint action; i.e. the dual of the adjoint action. The latter is the linearization at the identity of the action of G on G by conjugation: $h \mapsto ghg^{-1}$. The orbits of the coadjoint action are always symplectic manifolds. (This can be seen directly or by reducing T^*G by the action of G by left translation). Define $\{\{F, G\}\}$ at μ by restricting F and G to $\mathcal{O}_\mu$, the orbit of μ and then computing the symplectic Poisson bracket and evaluating at μ. Again this gives the same KAKS bracket.

We note that $\mathfrak{G}^*$ itself need not be a symplectic manifold (it may be odd-dimensional) but it is foliated by symplectic manifolds (route 2) and is closely associated to the symplectic manifold T^*G (route 1).

An important property of KAKS systems is that they leave the coadjoint orbits invariant. This is implicit in route 2 above, but may also be checked directly. This fact is related to a conservation law in the sense of Noether reflecting the G-symmetry of the problem.

If G has a biinvariant inner product $\langle , \rangle$ then the KAKS equations can be given a little more explicitly. In this case we can identify $\mathfrak{G}$ and $\mathfrak{G}^*$ via the Killing form $\langle , \rangle$ and the evolution equations on $\mathfrak{G}^*$ become evolution equations on $\mathfrak{G}$. (It is *not* always advantageous to make this identification). Denoting elements of $\mathfrak{G}$ by ξ, $\delta H/\delta \xi$ becomes $\nabla_\xi H$, the ξ-gradient of H, a vector field on $\mathfrak{G}$ and the KAKS equations are easily seen to be equivalent to the following equations on $\mathfrak{G}$:

$$\dot{\xi} = [\xi, \nabla_\xi H]$$

Example 1 (The Rigid Body). Here we take $G = SO(3)$ so that $\mathfrak{G}$, its Lie algebra, is identifiable with $\mathbb{R}^3$ and the Lie bracket with the cross product. A point $m \in \mathfrak{G}^*$ represents the angular momentum in "body coordinates". (See Abraham and Marsden [1978] or Arnold [1978] for the explanation of this terminology). For the moment we do not identify $\mathfrak{G}$ and $\mathfrak{G}^*$. The energy H is the kinetic energy of the body, a positive definite quadratic function of m. By choosing an appropriate orthonormal basis of $\mathbb{R}^3$ (and corresponding orthonormal dual basis of $\mathbb{R}^{3*}$) we can assume H is diagonal:

$$H(m) = \frac{1}{2}\left(\frac{m_1^2}{I_1} + \frac{m_2^2}{I_2} + \frac{m_3^2}{I_3}\right)$$

where I_1, I_2, I_3 are positive constants, the moments of inertia. Let us work out the KAKS equations $\dot{F} = \{\{F, H\}\}$ in this case. Clearly $\delta F/\delta m$ is just the vector in $\mathbb{R}^3$ with components $(\partial F/\partial m_1, \partial F/\partial m_2, \partial F/\partial m_3)$. Thus

$$\{\{F, H\}\}(m) = -\left\langle m, \frac{\delta F}{\delta m} \times \frac{\delta H}{\delta m}\right\rangle,$$

the triple product. Choosing $F(m) = m_1$, the equation $\dot{F} = \{\{F, H\}\}$ reads

$$\dot{m}_1 = -\begin{vmatrix} m_1 & m_2 & m_3 \\ 1 & 0 & 0 \\ \dfrac{m_1}{I_1} & \dfrac{m_2}{I_2} & \dfrac{m_3}{I_3} \end{vmatrix} = \frac{I_2 - I_3}{I_2 I_3} m_2 m_3$$

The equations for $\dot{m}_2$ and $\dot{m}_3$ are obtained by cyclic permutation. These are the famous Euler equations for a force-free rigid body. It is trivial to check that $(d/dt)(m_1^2 + m_2^2 + m_3^2) = 0$ i.e. $\|m\|^2$ is constant in time. The spheres $\|m\| =$ constant are exactly the coadjoint orbits for $SO(3)$. Thus $SO(3)^*$ is the union of these symplectic manifolds (plus the origin). Their preservation by the Euler equations corresponds to the conservation of angular momentum. There is a Killing form in the case, namely the standard inner product and $\dot{\xi} = [\xi, \nabla_\xi H]$ yields the same Euler equations.

EXAMPLE 2 (The Heavy Top). The equations for a heavy top can also be written in KAKS form for the Euclidean group. Here $E(3)=SO(3)\times\mathbb{R}^3$, the semi-direct product of rotations and translations. Writing elements of the Lie algebra as paris (m, v), the Hamiltonian is

$$H=\frac{1}{2}\left(\frac{m_1^2}{I_1}+\frac{m_2^2}{I_2}+\frac{m_3^2}{I_3}\right)+mgv_3$$

One can easily check that the KAKS equations yield the correct equations for a heavy top (see Guillemin and Sternberg [1980] and Ratiu (1981]). In case $I_1=I_2$ one has the Lagrange or symmetric top which has an additional S^1 symmetry. Using these geometric ideas, Holmes and Marsden [1981] showed that for certain $I_1\neq I_2$, the dynamics of the heavy top is chaotic; i.e. contains irregular aperiodic orbits whose closure is a complex invariant set. Here is a good example where traditional methods of classical applied mathematics need geometric augmentation in the spirit of Poincaré.

We refer to Arnold [1966], Abraham and Marsden [1978], Marsden, Ebin and Fischer [1972] and Marsden [1981] for the indication of how to similarly treat systems 3–6. In §3 we shall discuss 7 and 10. We now indicate how to treat 8 and 9 as they are amusing and perhaps not so well known.

EXAMPLE 8 (Quantum Mechanics). Here $\mathcal{U}(\mathcal{H})$ is the unitary group of a complex Hilbert space $\mathcal{H}$. The Schrodinger equation

$$i\dot{\psi}=H_{\mathrm{op}}\psi$$

where H_{op} is a self adjoint operator is well-known to be Hamiltonian relative to the symplectic form $\omega(\psi, \phi)=-\mathrm{Im}\langle\psi, \phi\rangle$ and the energy $H(\psi)=\frac{1}{2}\langle\psi, H_{\mathrm{op}}\psi\rangle$ (see Marsden [1968] or Chernoff and Marsden [1974] for instance).

The Heisenberg equations, equivalent to the Schrodinger equation, are:

$$i\dot{T}=[T, H_{\mathrm{op}}]$$

where T is a self-adjoint operator and [,] is the commutator bracket. These are in KAKS form. Indeed, the Lie algebra $\mathfrak{u}$ of $\mathcal{U}(\mathcal{H})$ consists of the skew adjoint operators; via the Hermitian inner product $\langle A, B\rangle=\mathrm{trace}(AB^*)$ we identify $\mathfrak{u}$ and $\mathfrak{u}^*$. The Hamiltonian is $H(A)=\langle iH_{\mathrm{op}}, A\rangle$, which is *real* and the KAKS bracket is

$$\begin{aligned}\{\{F, H\}\}(A)&=-\left\langle A, \left[\frac{\delta F}{\delta A}, iH_{\mathrm{op}}\right]\right\rangle\\&=\left\langle\frac{\delta F}{\delta A}, [iH_{\mathrm{op}}, A]\right\rangle\end{aligned}$$

Thus, either directly or by our general remarks, the KAKS equations are equivalent to

$$\dot{A} = i[H_{\rm op}, A]$$

But these are the Heisenberg equations with $A = iT$. The coadjoint orbits are the similarity classes $\{UAU^{-1} \mid u \in \mathcal{U}(\mathcal{H})\}$ and the fact that the evolution preserves these orbits is of course standard. (If we choose A to be $i \times$ projection onto a 1-dimensional subspace, its coadjoint orbit is just projective Hilbert space; the other orbits are also interesting symplectic manifolds).

Example 9. One can regard many Lax type equations that occur in non-linear wave equations (Lax [1968]) as KAKS equations in a similar way. (For the specific example of the Cologero system, which can be dealt with by the more general reduction methods, see Marsden [1981, p. 40ff] and for the KdV equation, see Abraham and Marsden [1978, §5.5])

Let $\mathcal{H}$ be a real or complex Hilbert space (finite or infinite dimensional), $\mathcal{U}$ the unitary group and $\mathfrak{u}$ its Lie algebra, the skew Hermitian matrices. Let $L: \mathfrak{u} \to \mathfrak{u}$ be a given (non-linear) operator and consider the "Lax equation"

$$\dot{\lambda} = [L(\lambda), \lambda]$$

for $\lambda \in \mathfrak{u}$. This equation means that λ evolves on a coadjoint orbit in $\mathfrak{u} \approx \mathfrak{u}^*$. Thus

$$\lambda(t) = U(t)\lambda(0)U(t)^{-1}$$

for a unitary operator $U(t)$, so $\lambda(0)$ and $\lambda(t)$ are unitarily equivalent (i.e. on the same coadjoint orbit) and so the evolution is isospectral. It is easily checked, as in the previous example, that the equations $\dot{\lambda} = [L(\lambda), \lambda]$ are in KAKS form if and only if

$$L(\lambda) = \frac{\delta H}{\delta \lambda}$$

for some function $H: \mathfrak{u} \to \mathbb{R}$ (with $\mathfrak{u}$ and $\mathfrak{u}^*$ identified by the Killing form $\langle A, B \rangle = \operatorname{trace}(AB^*)$.) Thus the equations are Hamiltonian as a KAKS system or in the standard sense on each coadjoint orbit in $\mathfrak{u}$.

§3. **The Poisson–Vlasov equation as a KAKS system.** The Poisson–Vlasov equation describes the evolution of a collisionless ionized plasma moving under the influence of self-induced electrostatic forces. When the electromagnetic field is dynamic, things are more interesting and these Maxwell–Vlasov equations are discussed later. The Poisson–Vlasov equation for a plasma moving in $\mathbb{R}^3$ is

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left(v^i \cdot \frac{\partial f}{\partial x^i} - \frac{1}{m} \frac{\partial \phi_f}{\partial x^i} \frac{\partial f}{\partial v^i} \right) = 0$$

where ϕ_f is defined by Poisson's equation:

$$\Delta\phi_f = -\rho_f$$

and ρ_f is the charge determined by f:

$$\rho_f(x, t) = e\int f(x, v, t)\, dv$$

Here $f(x, v, t)$ represents the plasma density in position-velocity space at time t; e is the ion charge and m the ion mass. We are considering the motion of a cloud of charged ions of a single species for simplicity. The generalization to several species is routine. For $e = 0$ we get Liouville's equation for free particles in $\mathbb{R}^3$.

The Hamiltonian for the Poisson–Vlasov equation is

$$H(f) = \frac{m}{2}\int \|v\|^2 f(x, v, t)\, dx\, dv + \frac{1}{2}\int \|\nabla\phi_f(x)\|^2\, dx$$

The Poisson–Vlasov equations can be recast in bracket form as

$$\dot{F} = \{\{F, H\}\}$$

where

$$\{\{F, G\}\}(f) = \int f\left\{\frac{\delta F}{\delta f}, \frac{\delta G}{\delta f}\right\} dx\, dv$$

and $\{,\}$ is the standard Poisson bracket for functions of x, v and $\delta F/\delta f$ here coincides with the functional derivative as used in physics and the calculus of variations.

The bracket form of the Poisson–Vlasov equations can be checked directly but is unsatisfactory unless we understand how this bracket is related to symplectic geometry. We shall now show that in fact the bracket is a KAKS bracket for a Lie group. In what follows we choose natural units in which $e = m = 1$. Moreover, we shall identify velocity with momentum; hence we let $\mathbb{R}^6$ denote the usual position-momentum phase space with coordinates $(x^1, x^2, x^3, p_1, p_2, p_3)$ and the standard symplectic structure $\omega = \sum dx^i \wedge dp_i$. Let $\mathcal{S}$ denote the group of canonical transformations of $\mathbb{R}^6$ (which satisfy certain growth conditions at infinity). The Lie algebra $\mathfrak{s}$ of $\mathcal{S}$ consists of the Hamiltonian vector fields on $\mathbb{R}^6$ (again with certain growth conditions). We shall identify elements of $\mathfrak{s}$ with their generating functions, so that $\mathfrak{s}$ consists of the C^∞ functions on $\mathbb{R}^6$ and the (right) Lie algebra structure is given by $[f, g] = -\{f, g\}$, the *negative* of the usual Poisson bracket on phase space. (This follows from Exercise 4.1G and Corollary 3.3.18 of Abraham and Marsden [1978]).

The dual space $\mathfrak{s}^*$ can be identified with the distribution densities on $\mathbb{R}^6$ (which satisfy certain decay conditions at infinity): the pairing between $h \in \mathfrak{s}$

and $f \in \mathfrak{s}^*$ is given by integration

$$\langle h, f\rangle = \int hf\,dx\,dp.$$

(The "density" is really $f\,dx\,dp$, but we denote it simply by f.) Now as for any Lie algebra, the dual space $\mathfrak{s}^*$ carries a natural Poisson structure.

The general formula $\{\{F, G\}\}(\mu) = -\left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu}\right]\right\rangle$ becomes, in this case,

$$\{\{F, G\}\}(f) = \int f \left\{\frac{\delta F}{\delta f}, \frac{\delta G}{\delta f}\right\} dx\,dp$$

which is the bracket mentioned above. Thus indeed the Poisson–Vlasov equations can be put exactly into KAKS form.

The preservation of coadjoint orbits is equivalent to the fact that f at time t is related to f at time $t = 0$ by composition with some canonical transformation. This is an important property of the Poisson–Vlasov equation. It can also be seen by writing the equation as a non-linear self-consistent Liouville system:

$$\frac{\partial f}{\partial t} = -\{f, \mathcal{H}(f)\}$$

where $\{,\}$ is the ordinary Poisson bracket in xp space and

$$\mathcal{H}(f) = \tfrac{1}{2}\|v\|^2 + \phi_f(x)$$

§4. **Maxwell's equations.** Before coupling the Vlasov equation to the electromagnetic field equations, we shall consider separately the Hamiltonian description of Maxwell's equations. The appropriate Poisson bracket for the electric (E) and magnetic (B) fields will be constructed by reduction (Marsden and Weinstein [1974]).

As the configuration space for Maxwell's equations, we take the space $\mathfrak{A}$ of vector fields A on $\mathbb{R}^3$. (These are the "vector potentials. In more general situations, one should replace $\mathfrak{A}$ by the set of connections on a principal bundle over configuration space.) The corresponding phase space is then the cotangent bundle $T^*\mathfrak{A}$. Elements of $T^*\mathfrak{A}$ may be identified with pairs (A, Y), where Y is a vector field density on $\mathbb{R}^3$. (As usual, we do not distinguish Y and $Y\,dx$.) The pairing between A's and Y's is given by integration, so that the canonical symplectic structure ω on $T^*\mathfrak{A}$ is given by

$$\omega((A_1, Y_1), (A_2, Y_2)) = \int_M (Y_2 \cdot A_1 - Y_1 \cdot A_2)\,dx$$

with associated Poisson bracket

$$\{F, G\} = \int \left(\frac{\delta F}{\delta A}\frac{\delta G}{\delta Y} - \frac{\delta F}{\delta Y}\frac{\delta G}{\delta A}\right) dx$$

With the Hamiltonian

$$H(A, Y)=\frac{1}{2}\int |Y|^2\, dx+\frac{1}{2}\int |\mathrm{curl}\, A|^2\, dx$$

Hamilton's equations are easily computed to be

$$\frac{\partial Y}{\partial t}=-\mathrm{curl}\,\mathrm{curl}\, A \quad \text{and} \quad \frac{\partial A}{\partial t}=Y$$

If we write B for curl A and E for $-Y$, the Hamiltonian becomes the usual field energy

$$\frac{1}{2}\int |E|^2\, dx+\frac{1}{2}\int |B|^2\, dx$$

and so Hamilton's equations imply Maxwell's equations:

$$\frac{\partial E}{\partial t}=\mathrm{curl}\, B \quad \text{and} \quad \frac{\partial B}{\partial t}=-\mathrm{curl}\, E$$

The remaining two Maxwell equations will appear as a consequence of gauge invariance. The gauge group $\mathcal{G}$ consists of real valued functions on $\mathbb{R}^3$; the group operation is addition. An element $\psi \in G$ acts on $\mathfrak{A}$ by the rule(1)

$$A \mapsto A+\nabla\psi.$$

This "translation" of A extends in the usual way to a canonical transformation ("extended point transformation") of $T^*\mathfrak{A}$ given by

$$(A, Y) \mapsto (A+\nabla\psi, Y).$$

Notice that the Hamiltonian $H(A, Y)$ is invariant under these transformations. This means that we can use the gauge symmetries to reduce the degrees of freedom of our system. The action of $\mathcal{G}$ on $T^*\mathfrak{A}$ has a corresponding conserved quantity, namely a map $J: T^*\mathfrak{A} \to \mathfrak{G}^*$ where $\mathfrak{G}$, the Lie algebra of $\mathcal{G}$, is identified with the real valued functions on $\mathbb{R}^3$. The map J, called a momentum map, may be determined by a standard formula (Abraham and Marsden [1978, Corollary 4.2.11]): for $\phi \in \mathfrak{G}$,

$$\langle J(A, Y), \phi\rangle=\int (Y \cdot \nabla\phi)\, dx=-\int (\mathrm{div}\, Y)\phi\, dx$$

Thus we may write

$$J(A, Y)=-\mathrm{div}\, Y.$$

---

(1) Notice that we work directly with three dimensional fields. Four dimensionally, one has an extra "degree" of gauge freedom associated with the time derivative $\partial_t\phi$. We have already eliminated this freedom and the corresponding non-dynamical field A_4 (whose conjugate momentum vanishes). This is the standard Dirac procedure for a relativistic field theory such as Maxwell's equations.

If ρ is an element of $\mathfrak{G}^*$ (i.e. ρ is a density on $\mathbb{R}^3$), $J^{-1}(\rho) = \{(A, Y) \in T^*\mathfrak{A} \mid \operatorname{div} Y = -\rho\}$. In terms of E, the condition $\operatorname{div} Y = -\rho$ becomes the Maxwell equation $\operatorname{div} E = \rho$, so we may interpret the elements of $\mathfrak{A}^*$ as charge densities.

By a general theorem on reduction (Marsden and Weinstein (1974]), the manifold $J^{-1}(\rho)/G$ has a naturally induced symplectic structure.

A little computation shows that the reduced manifold $J^{-1}(\rho)/\mathcal{G}$ can be identified with $\text{Max} = \{(E, B) \mid \operatorname{div} E = \rho, \operatorname{div} B = 0\}$, and that the Poisson bracket induced on Max is given in terms of E and B by

$$\{F, G\} = \int \left(\frac{\delta F}{\delta E} \operatorname{curl} \frac{\delta G}{\delta B} - \frac{\delta G}{\delta E} \operatorname{curl} \frac{\delta F}{\delta B} \right) dx$$

Maxwell's equations with an ambient charge density ρ are thus Hamilton's equations for

$$H(E, B) = \frac{1}{2} \int (|E|^2 + |B|^2)\, dx$$

on the space Max, and can be written

$$\dot{F} = \{F, H\}.$$

§5. **The Maxwell–Vlasov equations.** The *Maxwell–Vlasov* equations are:

$$\frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + \frac{e}{m}\left(E + \frac{v \times B}{c}\right) \cdot \frac{\partial f}{\partial v} = 0$$

$$\frac{1}{c}\frac{\partial B}{\partial t} = -\operatorname{curl} E$$

$$\frac{1}{c}\frac{\partial E}{\partial t} = \operatorname{curl} B - \frac{e}{c}\int v f(x, v, t)\, dv$$

$$\operatorname{div} E = \rho_f, \quad \text{where} \quad \rho_f = e\int f(x, v, t)\, dv$$

$$\operatorname{div} B = 0$$

(Letting $c \to \infty$ leads to the *Poisson–Vlasov equation*).

The Hamiltonian for the Maxwell–Vlasov system is

$$H(f, E, B) = \int \tfrac{1}{2} |v|^2 f(x, v, t)\, dx\, dv + \int \tfrac{1}{2}[|E(x, t)|^2 + |B(x, t)|^2]\, dx$$

Our goal is to understand the Hamiltonian structure of these equations. (As usual, we let $e = m = c = 1$).

The Hamiltonian structure for the Maxwell–Vlasov system is very simple if we choose as our variables densities on (x, p) space (rather than (x, v) space) and elements (A, Y) of $T^*\mathfrak{A}$. To avoid confusion with densities f on (x, v) space, we shall use the notation f_{mom} for densities on (x, p) space.

The Poisson structure on $\mathfrak{g}^* \times T^*\mathfrak{A}$ is just the sum of those on $\mathfrak{g}^*$ and $T^*\mathfrak{A}$: for functions $\bar{F}$ and $\bar{G}$ of f_{mom}, A, and Y, set

$$\{\{\bar{F}, \bar{G}\}\}(f_{\text{mom}}, A, Y) = \int f_{\text{mom}} \left\{\frac{\partial \bar{F}}{\delta f_{\text{mom}}}, \frac{\delta \bar{G}}{\delta f_{\text{mom}}}\right\} dx\, dp + \int \left(\frac{\delta \bar{F}}{\delta A}\frac{\delta \bar{G}}{\delta Y} - \frac{\delta \bar{G}}{\delta A}\frac{\delta \bar{F}}{\delta Y}\right) dx$$

and the Hamiltonian is just $H(f, E, B)$ written in terms of these variables. Using the classical relation $p = v + A$ between momentum and velocity we get

$$\bar{H}(f_{\text{mom}}, A, Y) = \frac{1}{2}\int |p - A(x)|^2 f_{\text{mom}}(x, p)\, dx\, dp + \frac{1}{2}\int (|Y|^2 + |\text{curl } A|^2)\, dx$$

One now computes easily that the evolution equations $\dot{\bar{F}} = \{\{\bar{F}, \bar{H}\}\}$ for a function $\bar{F}$ on $\mathfrak{g}^* \times T^*\mathfrak{A}$ are the Maxwell–Vlasov equations with $\dfrac{\partial B}{\partial t} = -\text{curl } E$ replaced by $\dfrac{\partial A}{\partial t} = Y$. To get the Maxwell–Vlasov equations as written we must, as in the case of Maxwell's equations, reduce by the action of $\mathcal{G}$.

The natural action of $\mathcal{G}$ on $\mathfrak{g}^*$ is defined by letting $\psi \in \mathcal{G}$ act by the (linear) map

$$f_{\text{mom}} \mapsto f_{\text{mom}} \circ \tau_{-\nabla\psi}$$

where $\tau_{-\nabla\psi} : \mathbb{R}^6 \to \mathbb{R}^6$ is the "momentum translation map" defined by

$$\tau_{-\nabla\psi}(x, p) = (x, p - \nabla\psi(x)).$$

It is easy to verify that $\tau_{-\nabla\psi}$ is a canonical transformation, so it preserves the ordinary Poisson bracket on $\mathbb{R}^6$. It follows that this action preserves the Poisson structure on $\mathfrak{g}^*$. A simple calculation shows that it has a momentum map $J : \mathfrak{g}^* \to \mathfrak{G}^*$ given by $J(f_{\text{mom}}) = -\int f_{\text{mom}}(x, p)\, dp$.

Now we define the action of $\mathcal{G}$ on the product $\mathfrak{g}^* \times T^*\mathfrak{A}$ as follows: $\psi \in \mathcal{G}$ maps

$$(f_{\text{mom}}, A, Y) \mapsto (f_{\text{mom}} \circ \tau_{-\nabla\psi}, A + \nabla\psi, Y).$$

This action leaves the Hamiltonian $\bar{H}$ invariant.

The momentum map $J : \mathfrak{g}^* \times T^*\mathfrak{A}\mathfrak{G}^*$ for this action is given by:

$$J(f_{\text{mom}}, A, Y) = -\int f_{\text{mom}}(x, p)\, dp - \text{div } Y.$$

With this action we are ready to consider the reduced manifold $(\mathfrak{g}^* \times T^*\mathfrak{A})_0 = J^{-1}(0)/\mathcal{G}$. This space may be identified with the *Maxwell–Vlasov phase space*

$$\mathcal{M}\mathcal{V} = \left\{(f, B, E) \mid \text{div } B = 0 \quad \text{and} \quad \text{div } E = \int f(x, v)\, dv\right\}$$

by associating to each (f_{mom}, A, Y) in $J^{-1}(0)$ the triple (f, B, E) in $\mathcal{MV}$ where

$$f(x, v) = f_{\mathrm{mom}}(x, v + A(x)),\ B = \operatorname{curl} A, \quad \text{and} \quad E = -Y.$$

By the general theory of reduction, $\mathcal{MV}$ inherits a Poisson structure from the one on $\mathfrak{g}^* \times T^*\mathfrak{A}$. Since the Hamiltonian $(\bar{H})$ is invariant under $\mathcal{G}$, it follows that the Maxwell–Vlasov equations are a Hamiltonian system on $\mathcal{MV}$ with respect to this structure. One can, in fact with a little labor compute the explicit form of the inherited Poisson structure in the variables (f, B, E). It is given by the (perhaps unexpectedly complex) formula

$$\begin{aligned}\{\{F, G\}\}(f, E, B) = &\int f\left\{\frac{\delta F}{\delta f}, \frac{\delta G}{\delta f}\right\} dx\, dv + \int\left(\frac{\delta F}{\delta E} \operatorname{curl} \frac{\delta G}{\delta B} - \frac{\delta G}{\delta E} \operatorname{curl} \frac{\delta F}{\delta B}\right) dx\\ &+ \int\left(\frac{\delta F}{\delta E} \cdot \frac{\partial f}{\partial v}\frac{\delta G}{\delta f} - \frac{\delta G}{\delta E} \cdot \frac{\partial f}{\partial v}\frac{\delta F}{\delta f}\right) dx\, dv\\ &+ \int fB \cdot \left(\frac{\partial}{\partial v}\frac{\delta F}{\delta f} \times \frac{\partial}{\partial v}\frac{\delta G}{\delta f}\right) dx\, dv\end{aligned}$$

The Maxwell–Vlasov equations can thus be written as

$$\dot{F} = \{\{F, H\}\}$$

We know by general principles that the bracket $\{\{,\}\}$ must satisfy Jacobi's identity. Indeed we obtained it by totally natural constructions. To verify Jacobi's identity by hand would be extremely tedious. In fact a first attempt at producing such a bracket was done by hand using brilliant guesswork by Morrison [1980]. However his bracket fails to satisfy Jacobi's identity (Weinstein and Morrison [1981]).

Having a Hamiltonian structure for the Maxwell–Vlasov equations is of great interest to plasma physicists because

(a) it enables them to begin to make use of powerful techniques of perturbation theory for Hamiltonian systems

(b) the relationship between the classical, quantum and semiclassical theories can be attacked in a useful context.

(c) the relationship between the full theory and various Hamiltonian truncations (such as the three wave interaction model) should now be understandable.

Similar Hamiltonian structures to these that we have found for plasmas can also be given for the equations of magnetohydrodynamics; cf. Morrison and Greene [1980].

Hopefully the situation described above reinforces the fact that current geometric methods in mechanics do have a role to play in concrete physical problems.

REFERENCES

R. Abraham and J. Marsden [1978]. *Foundations of Mechanics*, Second Edition, W. A. Benjamin Co.

V. Arnold [1966]. Sur la geometrié differentielle des groupes de Lie de dimension infinite et ses applications a l'hydrodynamique des fluids parfaits, *Ann. Inst. Grenoble* **16**, 319–361.

V. Arnold [1968]. *Mathematical methods of classical mechanics*. Springer Graduate Texts in Math. No. 60 Springer-Verlag, New York.

P. Chernoff and J. Marsden [1974]. *Properties of Infinite Dimensional Hamiltonian Systems*, Springer Lecture Notes #425.

A. J. Chorin, T. J. R. Hughes, M. F. McCracken and J. E. Marsden [1978]. Product formulas and numerical algorithms, *Comm. Pure and Appl. Math.* **31**, 205–256.

D. Ebin and J. Marsden [1970]. Groups of diffeomorphisms and the motion of an incompressible fluid, *Ann. of Math.* **92**, 102–163.

V. Guillemin and S. Sternberg [1980]. The moment map and collective motion, *Ann. of Phys.* **127**, 220–253.

P. J. Holmes and J. E. Marsden [1981]. Horseshoes and Arnold Diffusion for Hamiltonian systems on Lie Groups (to appear).

P. D. Lax [1968]. Integrals of nonlinear equations of evolution and solitary waves, *Comm. Pure and Appl. Math.* **21**, 467–490.

J. E. Marsden [1968]. Hamiltonain One-Parameter Groups, *Arch. Rat. Mech. An.* **28**, 362–396.

J. E. Marsden [1981]. *Lectures on geometric methods in mathematical physics*, CBMS-NSF Conference Series #37, SIAM.

J. E. Marsden, D. G. Ebin and A. E. Fischer [1972]. Diffeomorphism Groups, Hydrodynamics and General Relativity, *13th Biennial Seminar of Can. Math. Congress* (J. Vanstone, ed.) 135–279.

J. E. Marsden and A. Weinstein [1974]. Reduction of symplectic manifolds with symmetry, *Reports on Math. Phys.* **5**, 121–130.

J. Marsden and A. Weinstein [1982]. The Hamiltonian Structure of the Maxwell–Vlasov equations, *Physica D* (to appear).

P. J. Morrison [1980]. The Maxwell–Vlasov equations as a continuous Hamiltonian system, *Phys. Lett.* **80A**, 383–386.

P. J. Morrison and J. M. Greene [1980]. Noncanonical Hamiltonian density formulation of hydrodynamics and ideal magnetohydrodynamics, *phys. Rev. Letters.* **45**, 790–794.

T. Ratiu [1981]. Euler–Poisson equations on Lie algebras and the N-dimensional heavy rigid body, *Proc. Natl. Acad. Sci.* U.S.A. **78**, 1327–1328.

A. Weinstein and P. Morrison [1981]. Comments on: The Maxwell–Vlasov equations as a continuous Hamiltonian system, *Phys. Lett.* **86A**, 235–236.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

Canad. Math. Bull. Vol. **25** (2), 1982

A NOTE ON COMPLETELY AND ABSOLUTELY MONOTONE FUNCTIONS

BY
ARVIND MAHAJAN AND DIETER K. ROSS

ABSTRACT. The solutions of a certain class of first order linear differential equations are shown to be either completely or absolutely monotone depending on the nature of its coefficients. This is a simple theorem which is used to deduce a number of new and interesting results dealing with the complete and absolute monotonicity of functions. In particular, a partial answer is supplied to a question posed by Askey and Pollard: "When is $x^{-c_0}(1^2+x^2)^{-c_1}\cdots(n^2+x^2)^{-c_n}$ completely monotone?"

1. **Introduction.** During the past decade it has become apparent that the notions of completely and absolutely monotone function play an important role in classical analysis. They occur, for example, in proving the positivity of integrals involving Bessel functions or the positivity of Cesàro means of certain Jacobi series; see Askey [1]. The notion of complete monotonicity has also been used in characterising the gamma function [6].

In the past various methods have been developed to prove that particular functions are either completely or absolutely monotone; see for example the work of Askey and Pollard [2] and of Fields and Ismail [4]. The purpose of this note is to show that some additional and some known results follow directly from a simple theorem dealing with the nature of the solutions to first order linear differential equations. In particular, it is shown that $x^{-c_0}(1^2+x^2)^{-c_1}\cdots(n^2+x^2)^{-c_n}$ is completely monotone for certain values of c_i.

To begin, recall that a function $f(x)$ is said to be k-times monotone on $(0,\infty)$ if all its derivatives $f^{(i)}(x)$ exist and satisfy

$$(1)\qquad (-1)^i f^{(i)}(x)\geq 0 \qquad (0<x<\infty;\ i=0,1,2,\ldots,k).$$

In case k is infinite, $f(x)$ is said to be completely monotone on $(0,\infty)$. On the other hand, a function $f(x)$ is said to be absolutely monotone on $(0,\infty)$ if (1)

Received by the editors September 27, 1979 and, in revised form, April 8, 1980, August 14, 1980 and November 26, 1980.

AMS subject classifications (1980). Primary 26D15, 34A10; Secondary 33A15, 33A40, 40G05.

Key words and phrases. Completely monotone, absolutely monotone, Inequalities, Bessel functions, gamma functions.

can be replaced by

$$f^{(i)}(x)\geq 0 \qquad (0<x<\infty;\ i=0,1,2,\ldots).$$

An important and fundamental result which will be needed here is the so-called Bernstein-Widder theorem which states that a necessary and sufficient condition for a function $f(x)$ to be completely monotone on $(0,\infty)$ is that it should be representable as a Laplace-Stieltjes type integral of the form

$$f(x)=\int_0^\infty e^{-xt}\,du(t),$$

where $du(t)\geq 0$, and the integral converges for $0<x<\infty$.

For further discussion on the properties and characteristics of completely and absolutely monotone functions, see Widder [8].

2. **Completely monotone functions.** Consider the following simple theorem:

THEOREM 1. *Let $p(x)$ be a completely monotone function defined on* $(0,\infty)$. *Then all solutions of the differential equation*

$$y'(x)+p(x)y(x)=0 \tag{2}$$

which are nonnegative for a single $x_0>0$ are completely monotone on $(0,\infty)$.

The proof of this theorem is straightforward and depends on the fact that the solutions in question can be written in the form

$$y(x)=y(x_0)\exp\left[\int_{x_0}^{x}-p(t)\,dt\right]\geq 0, \qquad 0<x<\infty.$$

The required result follows by differentiation of (2).

The above is an elementary theorem which can be used to obtain a number of known as well as new results on completely and absolutely monotone functions.

EXAMPLE 1. (The question posed by Askey and Pollard [2]). The function $y(x)=x^{-c_0}(1^2+x^2)^{-c_1}\cdots(n^2+x^2)^{-c_n}$ is completely monotone for $x>0$ provided that

$$c_0\geq 2c_1\geq 0 \quad\text{and}\quad 0\leq c_i\leq c_{i-1}\left[1-\frac{1}{i+\alpha}\right], \qquad (i=2,3,\ldots,n) \tag{3}$$

where $-1<\alpha\leq A$ and $A=4.5678018\cdots$ is the positive root of

$$9x^7+55x^6-14x^5-948x^4-3247x^3-5013x^2-3780x-1134=0.$$

Consider the logarithmic derivative of $y(x)$ as given by

$$\begin{aligned}\frac{-y'(x)}{y(x)}&=\frac{c_0}{x}+\frac{2c_1x}{(1+x^2)}+\cdots+\frac{2c_nx}{(n^2+x^2)}\\&=2\int_0^\infty e^{-xt}\left[\frac{c_0}{2}+c_1\cos t+\cdots+c_n\cos nt\right]dt\end{aligned} \tag{4}$$

then, because

$$a_0+a_1\cos t+\frac{a_2(1+\alpha)}{(2+\alpha)}\cos 2t+\cdots+\frac{a_n(1+\alpha)}{(n+\alpha)}\cos nt\geq 0$$

for $a_0\geq a_1\geq\cdots\geq a_n\geq 0$ and under the above conditions stated on α, see Gasper [5, Theorem 2] (where A has been replaced by α in order to correct a misprint), it follows that the right side of (4) is completely monotone. Hence, by Theorem 1 the function $y(x)$ is completely monotone.

EXAMPLE 2. The conditions referred to in (3) require that c_1 be positive. However, it is possible to include instances where $c_1<0$. To begin, consider the function $\phi(x)=x^{-c_0}(1^2+x^2)^{-c_1}$ with $c_0\geq 2|c_1|$ then

$$\frac{-\phi'(x)}{\phi(x)}=\int_0^\infty e^{-xt}[c_0+2c_1\cos t]\,dt$$

which, together with Theorem 1, implies that $\phi(x)$ is completely monotone on $(0,\infty)$ since the condition $c_0\geq 2|c_1|$ ensures that the integrand is positive. On using

$$x^{-1}(1^2+x^2)^{-1}=\int_0^\infty e^{-xt}[1-\cos t]\,dt$$

and the fact that a product of completely monotone functions is itself completely monotone, it appears that $x^{-c_0-m}(1+x^2)^{-c_1-m}$ is also completely monotone for $x>0$, $c_0\geq 2\,|c_1|$ and $m=0,1,\ldots$.

COROLLARIES TO EXAMPLE 2. (a) *The function* $x^{-2|\tau|}(1+x^2)^{-\tau}$ *is completely monotone for any* τ *real,* $x>0$. *This result was proved by Askey and Pollard* [2] *who used two different methods, the first making use of a theorem of I. J. Schoenberg and the second making use of an inductive argument. (Incidentally, these authors also point out that the function* $x^{-\kappa\tau}(1+x^2)^{-\tau}$ *cannot be completely monotone for all* $\tau>0$ *if* $\kappa<2$.)

A similar result, namely the complete monotonicity of $x^{-2\tau}(1+x^2)^{-\tau}$, when $\tau>0$ and $x>0$, was proved independently by Fields and Ismail [3]. In another paper by the same authors, [4], it is shown that the function $x^{2\tau-2\rho\tau-b}(1+x^2)^{-\tau}$ is completely monotone for $b\geq 0$, $\rho\geq 2$, $\tau>0$, $x>0$ and other more complicated results are deduced by using an argument of Darboux type and fractional integration. However, the result mentioned here is a trivial consequence of Corollary (a).

(b) *The function* $x^{-d-m}(1+x^2)^{-c-m}$ *is completely monotone for* $d=2c$, $0<c\leq\frac{1}{2}$, $m=0,1,\ldots,$ *and the function* $x^{-(2c-1)}(1+x^2)^{-c}$ *is completely monotone for* $c\geq 1$. *These results follow from Example* 2 *and also appear in Askey* [1].

The significance of the notion of complete monotonicity is that it can be combined with the earlier mentioned Bernstein-Widder theorem to verify that certain integrals involving special functions are positive. By way of example we follow Askey [1] and use the previous results. Thus, because

$$(1+x^2)^{-\alpha-1/2}=\frac{\Gamma(\frac{1}{2})}{2^\alpha\Gamma(\alpha+\frac{1}{2})}\int_0^\infty e^{-xt}t^\alpha J_\alpha(t)\,dt,\quad \text{for}\quad \alpha>-\tfrac{1}{2},\quad x>0,$$

see Watson [7, p. 386] it follows, on account of the convolution theorem for Laplace transforms, that

$$\int_0^\infty e^{-xt}\left(\int_0^t (t-u)^\beta u^\alpha J_\alpha(u)\,du\right)dt=\frac{2^\alpha\Gamma(\alpha+\frac{1}{2})\Gamma(\beta+1)}{\Gamma(\frac{1}{2})x^{\beta+1}(1+x^2)^{\alpha+\frac{1}{2}}}$$

when $\beta>-1$ and since $x^{-(\beta+1)}(1+x^2)^{-(\alpha+\frac{1}{2})}$ is completely monotone for $\beta\geq 2\alpha$ that

$$\int_0^x (x-t)^{2\alpha+\varepsilon}t^\alpha J_\alpha(t)\,dt\geq 0,$$

for $\alpha>-\frac{1}{2}$, $\varepsilon\geq 0$ and $x>0$.

In much the same way, by making use of the identity

$$x(1+x^2)^{-\alpha-3/2}=\frac{2^{-\alpha-1}\Gamma(\frac{1}{2})}{\Gamma(\alpha+3/2)}\int_0^\infty e^{-xt}t^{\alpha+1}J_\alpha(t)\,dt,\ \alpha>-1$$

Watson [7, p. 386], we may obtain the inequality

$$\int_0^x (x-t)^{2\alpha+2+\varepsilon}t^{\alpha+1}J_\alpha(t)\,dt\geq 0,$$

for $\alpha\geq-\frac{1}{2}$, $\varepsilon\geq 0$ and $x>0$.

For $\varepsilon=0$ these reduce to known results, see Fields and Ismail [4] and Askey [1].

EXAMPLE 3. During an investigation of the monotonicity of the function $\psi(x)=[x^\alpha\Gamma(x)(e/x)^x]^\beta$ on $(0,\infty)$ for various values of α and β, Muldoon [6, Theorem 2.1] uses Stirling's formula together with some properties of completely and absolutely monotone functions. The same result can be obtained more simply by using Theorem 1 and noting that

$$\begin{aligned}\frac{-\psi'(x)}{\psi(x)}&=-\beta\left[\alpha/x+\frac{\Gamma'(x)}{\Gamma(x)}-\log x\right]\\&=\beta\int_0^\infty e^{-xt}\left[\frac{1}{1-e^{-t}}-1/t-\alpha\right]dt\end{aligned}$$

(Binet's formula) and that the integrand is positive for $\alpha\leq\frac{1}{2}$ and where $\beta\geq 0$. In case $\alpha>\frac{1}{2}$ it is known from the asymptotic form of $\psi(x)$, see Muldoon [6], that this function is ultimately increasing and so cannot be completely monotone for all $x>0$.

3. **Absolutely monotone functions.** A corresponding theorem for absolutely monotone functions is the following:

THEOREM 2. *Let* $q(x)$ *be an absolutely monotone function defined on* $(0, \infty)$. *Then all the solutions of the differential equation*

$$y'(x) - q(x)y(x) = 0$$

which are nonnegative for a single $x_0 > 0$, *are absolutely monotone on* $(0, \infty)$.

EXAMPLE 4. Here we shall use the above theorem to reproduce a result of Askey and Pollard [2, Theorem 2], namely, that the function $\zeta(x) = (1-x)^{2\tau}(1-2xh+x^2)^{-\tau}$ is absolutely monotone for $\tau < 0$ and with $-1 \le h = \cos\theta \le 1$.

To see this, consider the logarithmic derivative of $\zeta(x)$, namely,

$$\frac{\zeta'(x)}{\zeta(x)} = \frac{-2\tau(1-h)(1+x)}{(1-x)(1-2xh+x^2)}$$

and since Askey and Pollard [2] have pointed out that the absolute monotonicity of $\dfrac{(1+x)}{(1-x)(1-2xh+x^2)}$ follows easily from Fejér's theorem on positivity of the $(C, 1)$ means for Fourier series, it follows that $\dfrac{\zeta'(x)}{\zeta(x)}$ is absolutely monotone and hence by Theorem 1 the required result follows.

We conclude this note by giving just one more theorem of this kind. Obviously, many such theorems can be obtained for first order differential equations.

THEOREM 3. *Consider the first order inhomogeneous linear differential equation*

$$y'(x) - f(x)y(x) = g(x) \tag{5}$$

where $f(x)$ *and* $g(x)$ *are continuous on* $x > 0$. *Suppose further that* $f(x)$ *and* $g(x)$ *are absolutely monotone for* $x > 0$. *Then* (5) *has a nontrivial solution which is also absolutely monotone on this interval.*

ACKNOWLEDGEMENTS. The authors wish to thank the referee for his comments. One of the authors (A.M.) is grateful to the Australian Government for a Postgraduate Research Award.

REFERENCES

1. R. Askey, *Summability of Jacobi series*, Trans. Amer. Math. Soc., **179** (1973), pp. 71–84.

2. R. Askey and H. Pollard, *Some absolutely monotonic and completely monotonic functions*, SIAM J. Math. Anal. **5** (1974), pp. 58–63.

3. J. L. Fields and M. Ismail, *On some conjectures of Askey concerning completely monotonic functions*, Spline Functions and Approximation Theory, edited by A. Meir and A. Sharma, ISNM Vol. 21, Birkhaüser Verlag, Basel, 1973, pp. 101–111.

4. J. L. Fields and M. E.-H. Ismail, *On the positivity of some* ${}_1F_2$*'s*, SIAM J. Math. Anal., **6** (1975), pp. 551–559.

5. G. Gasper, *Nonnegative sums of cosine, ultraspherical and Jacobi polynomials*, J. Math. Anal. Appl. **26** (1969), pp. 60–68.

6. M. E. Muldoon, *Some monotonicity properties and characterisations of the gamma function*, Aequationes Math. **18** (1978), pp. 54–63.

7. G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge Univ. Press, 1966.

8. D. V. Widder, *Laplace Transforms*, Princeton University Press, Princeton, N.J., 1946.

DEPARTMENT OF APPLIED MATHEMATICS
LA TROBE UNIVERSITY
BUNDOORA, 3083, AUSTRALIA

Canad. Math. Bull. Vol. **25** (2), 1982

FIVE MUTUALLY TANGENT SPHERES AND VARIOUS ASSOCIATED CONFIGURATIONS

BY

J. F. RIGBY

ABSTRACT. If five spheres $\sigma_0, \sigma_1, \ldots, \sigma_4$ touch each other externally and have radii in geometrical progression, there is a dilative rotation mapping $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ to $\sigma_1, \sigma_2, \sigma_3, \sigma_4$; the dilatation factor is shown to be negative. The ten points of contact of the spheres lie by fours on 15 circles, forming a $(15_4 10_6)$ configuration in inversive space. In the corresponding configuration in the inversive plane, the 15 circles meet again in 60 points, which lie by fours on 45 circles touching by threes at each of the 60 points, and forming a configuration isomorphic to that of 60 Pascal lines (associated with six points on a conic) meeting by fours at 45 points. The 45 circles arise from ten Money–Coutts configurations of nine anti-tangent cycles. Conjectures are made about other circles through the 60 points.

1. **Introduction.** If five spheres $\sigma_0, \sigma_1, \ldots, \sigma_4$ all touch each other externally and have radii $1, r, r^2, r^3, r^4$ in geometric progression, then [2, p. 119] r satisfies the equation

$$r^2-(1+\sqrt{2})r+1=0, \tag{1}$$

and there is a similarity mapping $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ to $\sigma_1, \sigma_2, \sigma_3, \sigma_4$. This similarity, since it is not an isometry, is a dilative rotation (i.e. a rotation about a line l followed by a dilatation with centre O on l) [3, p. 102]. Coxeter tacitly assumes in [2] that the dilatation factor is positive, so that "the points of contact of consecutive spheres lie on a concho-spiral" [2, p. 119], but one of his research students, Mrs Asia Weiss, has recently shown that it is negative [6]. We give here a simple alternative proof of this fact.

Associated with five tangent spheres are a (10_6) configuration of spheres and points and also a $(15_6 10_9)$ configuration. We show that the two configurations exist also in a more general form.

Both these configurations give rise to a $(15_4 10_6)$ configuration of circles and points in three dimensions. Such a configuration exists also in two dimensions and it contains ten (9_4) configurations. The author has shown [5] that a (9_4) configuration gives rise to a Money–Coutts configuration of nine anti-tangent cycles, and we deduce from this that the $(15_4 10_6)$ configuration gives rise to a $(45_4 60_3)$ configuration of 45 circles touching by threes at 60 points.

Received by the editors February 22, 1980 and, in revised form, August 12, 1980.

AMS classification 51M99.

Key words and phrases. Inversive plane. Configurations of spheres, circles and points. Pascal and Cremona configurations. Money–Coutts configuration. Loxodromic sequences of tangent spheres.

This last configuration is isomorphic to the configuration of 45 points of intersection of the 60 Pascal lines of six points on a conic. Other geometrical properties of the Pascal configuration suggest further fruitful investigations into the geometry of the circles and points. I am grateful to Professor Coxeter for reminding me about the Pascal configuration.

2. **The dilative rotation.** In the notation of section 1, take O as origin, l as the z-axis. The dilative rotation obtained by rotation about l through an angle $\theta+\pi$ followed by a negative dilatation with factor $-\lambda$ is equivalent to a rotation through an angle θ followed by a positive dilatation with factor λ followed by a reflection in the xy-plane. Suppose if possible that this dilative rotation maps $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ to $\sigma_1, \sigma_2, \sigma_3, \sigma_4$; then $\lambda = r$. Let us express the point (x, y, z) by $(x+iy, z)$ and write $e^{i\theta}=\kappa$. Denote the centre of σ_0 by (a, b), where we may suppose that a is real; then the centre of σ_i is $(ar^i\kappa^i, b(-r)^i)$ $(i=1,\ldots,4)$ and its radius is r^i. The conditions for σ_0 to touch $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ externally are

$$a^2(r\kappa-1)(r\bar{\kappa}-1)+b^2(r+1)^2=(r+1)^2,$$
$$a^2(r^2\kappa^2-1)(r^2\bar{\kappa}^2-1)+b^2(r^2-1)^2=(r^2+1)^2,$$
$$a^2(r^3\kappa^3-1)(r^3\bar{\kappa}^3-1)+b^2(r^3+1)^2=(r^3+1)^2,$$
$$a^2(r^4\kappa^4-1)(r^4\bar{\kappa}^4-1)+b^2(r^4-1)^2=(r^4+1)^2.$$

If these conditions are satisfied, we see by applying the dilative rotation that all the spheres are mutually tangent. Let us write $b/a=c$, $1/a=d$; then the above conditions become

$$(r^2+1)-r(\kappa+\bar{\kappa})+c^2(r+1)^2=d^2(r+1)^2, \tag{2}$$
$$(r^4+1)-r^2(\kappa^2+\bar{\kappa}^2)+c^2(r^2-1)^2=d^2(r^2+1)^2, \tag{3}$$
$$(r^6+1)-r^3(\kappa^3+\bar{\kappa}^3)+c^2(r^3+1)^2=d^2(r^3+1)^2, \tag{4}$$
$$(r^8+1)-r^4(\kappa^4+\bar{\kappa}^4)+c^2(r^4-1)^2=d^2(r^4+1)^2. \tag{5}$$

If we eliminate c^2 and d^2 from equations (2) and (4), by taking (4)$-$$(r^2-r+1)^2\times$(2), we obtain

$$(r^6+1)-(r^2-r+1)^2(r^2+1)-r^3(\kappa^3+\bar{\kappa}^3)+(r^2-r+1)^2r(\kappa+\bar{\kappa})=0.$$

Dividing by r, and writing $\kappa+\bar{\kappa}=k$, we have

$$2(r^4-2r^3+2r^2-2r+1)-r^2(k^3-3k)+(r^2-r+1)^2k=0. \tag{6}$$

From (1), $r^2-r+1=\sqrt{2}r$, and $r^4-2r^3+2r^2-2r+1=(r^2-r+1)^2-r^2=2r^2-r^2=r^2$, so (6) becomes

$$k^3-5k-2\equiv(k+2)(k^2-2k-1)=0$$

on dividing through by $-r^2$.

Now $k=2\cos\theta$; hence the solution $k=1+\sqrt{2}$ is impossible. Hence $k=-2$ or $k=1-\sqrt{2}$.

If $k=-2$ then $\kappa=-1$, and (2) and (3) give

$$(1+\sqrt{2})+2+(3+\sqrt{2})c^2=(3+\sqrt{2})d^2,$$
$$(1+2\sqrt{2})-2+(-1+2\sqrt{2})c^2=(3+2\sqrt{2})d^2,$$

whence $c^2=-1$ and $d^2=0$, which is impossible.

Hence $k=1-\sqrt{2}=\kappa+\bar{\kappa}$, and we easily check that the unique solution of equations $(2),\ldots,(5)$ is given by $c^2=3(3\sqrt{2}-2)/14$, $d^2=(3\sqrt{2}-2)/2$. Hence $a^2=(3\sqrt{2}+2)/7$, $b^2=3/7$. Also $2\cos\theta=\kappa+\bar{\kappa}=1-\sqrt{2}$, so $\theta=101°57'$. (If we return to the original point of view and take the dilatation factor to be $-r$, the angle of rotation will be $78°03'$ in the opposite direction.)

Is it possible to have a different arrangement of spheres with radii $1, r, \ldots, r^4$ giving a *positive* dilatation factor? No: since all five radii are known, once we have placed $\sigma_0,\ldots,\sigma_3$ all touching each other externally, there is only one possible position for σ_4 (since the second sphere touching $\sigma_0,\ldots,\sigma_3$ has radius r^{-1}).

3. **The points of contact of the spheres.** If we denote the dilative rotation by α, and denote the sphere $\sigma_0\alpha^r$ by σ_r $(r=\ldots,-2,-1,0,1,2,\ldots)$ we obtain an infinite sequence of spheres, as described in [2], such that any five consecutive spheres are mutually externally tangent. Denote the centre of σ_i by C_i, and the z-coordinate of C_i by $z_i=b(-r)^i$. When $j=1$ or 3, we see that

σ_i and σ_{i+j} touch externally,

σ_i and σ_{i+j} have radii r^i and r^{i+j},

$z_{i+j}=-r^jz_i$.

Hence the point of contact $P_{i,i+j}$ of σ_i and σ_{i+j} has z-coordinate 0. Thus all the points of contact $\ldots, P_{01}, P_{12}, P_{23}, P_{34},\ldots$ and $\ldots, P_{03}, P_{14}, P_{25},\ldots$ are coplanar. Also it is easy to see that $\ldots, P_{01}, P_{12}, P_{23},\ldots$ lie on an equiangular spiral, and $\ldots, P_{03}, P_{14},\ldots$ on a congruent spiral.

Since any infinite sequence of spheres (τ_r) $(r=\ldots,-2,-1,0,1,2,\ldots)$, such that any five consecutive spheres are mutually tangent, is inversively equivalent to the sequence (σ_r), it follows that the points of contact $\ldots, Q_{01}, Q_{12}, Q_{23}, Q_{34},\ldots$ and $\ldots, Q_{03}, Q_{14}, Q_{25},\ldots$ all lie on a sphere (or plane), where Q_{ij} denotes the point of contact of τ_i and τ_j. Coxeter has given two simpler proofs of this result [4], but it is a nice consequence of the fact that the dilatation factor in the above discussion is negative.

4. **A (10_6) configuration.** Let us now consider again just five mutually tangent spheres $\tau_0,\tau_1,\ldots,\tau_4$. We have seen that $Q_{01}, Q_{12}, Q_{23}, Q_{34}, Q_{03}, Q_{14}$ lie on a sphere; these six points are the points of contact of τ_1 and τ_3 with the

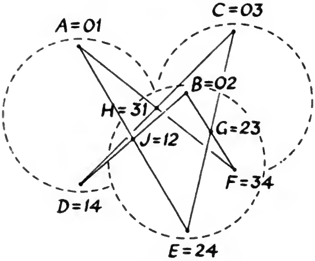

Figure 1

remaining spheres, so we can name the sphere on which they lie τ_{13} (or τ_{31}). The five mutually tangent spheres can be arranged in a sequence in 5! ways, but these sequences give us a total of only ten sets of six points of contact lying on ten spheres $\tau_{01}, \tau_{02}, \dots, \tau_{34}$.

If we invert Q_{04} to infinity, then τ_0 and τ_4 become parallel planes, and τ_1, τ_2, τ_3 become tangent spheres of equal radius sandwiched between the planes as in Fig. 1. The existence of the ten spheres, each containing six points of contact, is now obvious. A list of the spheres is given below. In Fig. 1 we have written $Q_{01} = A$, $Q_{02} = B$, etc., and $Q_{04} = K$ is at infinity.

$$\begin{array}{rr} BCDJHK = \tau_{01} & ACDFJG = \tau_{13} \\ CAEGJK = \tau_{02} & BCEFJH = \tau_{23} \\ ABFHGK = \tau_{03} & AEFJHK = \tau_{14} \\ ABCDEF = \tau_{04} & BFDGJK = \tau_{24} \\ ABDEHG = \tau_{12} & CDEHGK = \tau_{34} \end{array} \tag{7}$$

(Since K is at infinity in Fig. 1, six of the spheres appear as planes.) Each of the ten points lies on six spheres; thus we have a (10_6) configuration of spheres and points. By permuting the suffixes $0, 1, \dots, 4$ we obtain 5! symmetries of the configuration, and it is easily seen that there are no others: the symmetry group is S_5. Since we can permute the spheres $\tau_0, \tau_1, \dots, \tau_4$ in any manner using products of inversions, we can realize all the symmetries of the configuration within the inversive group.

5. **A $(15_6 10_9)$ configuration.** We see from Fig. 1 that there are five more spheres each containing six points of contact. One of them is $ABCGHJ$, containing the points of contact of $\tau_0, \tau_1, \tau_2, \tau_3$; we shall denote this sphere by

τ_{45}. The complete list of five spheres is

$$DEFGHJ = \tau_{05} \qquad ABDEJK = \tau_{35}$$
$$BCEFGK = \tau_{15} \qquad ABCGHJ = \tau_{45} \tag{8}$$
$$CAFDHK = \tau_{25}$$

If $ijklm$ is any permutation of 01234, we shall give Q_{ij} the alternative name $Q_{ij5,klm}$, where the suffix is an unordered pair of unordered triplets. Then we find that τ_{01} contains the points $Q_{023,145}$, $Q_{024,135}$, $Q_{025,134}$, $Q_{034,125}$, $Q_{035,124}$, $Q_{045,123}$, with similar results for the remaining fourteen spheres. Also $Q_{012,345}$ lies on the nine spheres $\tau_{03}, \tau_{04}, \tau_{05}, \tau_{13}, \tau_{14}, \tau_{15}, \tau_{23}, \tau_{24}, \tau_{25}$; etc.. Thus we have a $(15_6 10_9)$ configuration of spheres and points. By permuting the suffixes $0, 1, \ldots, 5$ we obtain 6! symmetries of the configuration, and it is easily seen that there are no others: the symmetry group is S_6.

Each of these two configurations contains fifteen circles with four points on each, each circle belonging to two spheres of the (10_6) configuration and to three spheres of the $(15_6 10_9)$ configuration. Using Fig. 1 we calculate that the three circles lying in one of the ten original spheres intersect at an angle of $2 \sin^{-1}(1/2\sqrt{2})$, whilst the three circles on the five other spheres intersect at right angles. Since inversions preserve angles, we can therefore realize only 5! of the symmetries of $(15_6 10_9)$ within the inversive group.

6. **The general $(15_6 10_9)$ configuration.** We have proved the existence of a special form of the $(15_6 10_9)$ configuration, derived from five mutually tangent spheres. We shall now investigate the most general manner of constructing the configuration. The general configuration is shown in Fig. 2, with K still at infinity. Let us analyse the figure to see how it can be constructed. The planes $BCEF$, $CAFD$, $ABDE$ meet at O, say; hence the lines AD, BE, CF meet at O. Denote the sphere $ABCDEF$ by σ. The points G, H, J lie in the polar plane π of O with respect to σ.

We begin therefore with a sphere σ, a point O not on σ, and two lines through O meeting σ at A, D and B, E; then J is determined. Now we require $BCJH$ to be concyclic; hence $DH.DC = DJ.DB = k_1$ say. Hence C lies on the inverse of π with respect to the sphere (real or imaginary) with centre D and radius $\sqrt{k_1}$. Similarly $EG.EC = EJ.EA = k_2$ say. Hence C lies on the inverse of π with respect to the sphere with centre E and radius $\sqrt{k_2}$. Thus C lies on two spheres as well as on σ. With suitable initial choices for σ, OAD and OBE, these three spheres will meet in two points, giving two possible positions for C; we choose either one of them. Now define $G = \pi \cap EC$, $H = \pi \cap DC$, $F = \sigma \cap OC$. To show that B, G, F are collinear, we observe that $BF \cap EC$ lies in π; but G is the point of EC in π, so $BF \cap EC = G$. Similarly H lies on AF.

Now $BCJH$ and $ACJG$ are concyclic; these circles have two points in

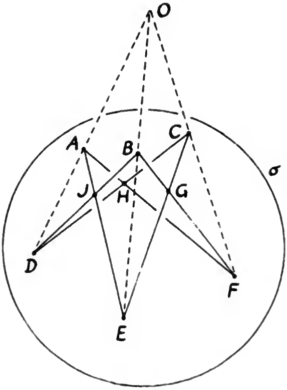

Figure 2

common, so *ABCGHJ* lie on a sphere. Hence *ABHG*, being coplanar also, are concyclic.

Now *BCEF* are concyclic (being coplanar and lying on σ) and *BCJH* are concyclic. Hence *BCEFJH* lie on a sphere. Similarly *CAFDGJ* and *ABDEHG* lie on spheres. From these spheres, cut by planes, we see that *FDGJ* and *DEHG* are concyclic. Hence *DEFGHJ* lie on a sphere.

We now have all the fifteen spheres, including the nine planes in the figure.

Since any configuration of five mutually tangent spheres can be mapped to any other by a product of inversions, we can say that such a configuration has no degrees of freedom *with respect to the inversive group.* It is easily seen that the general $(15_6 10_9)$ configuration has four degrees of freedom.

7. **The $(15_4 10_6)$ configuration in the plane.** If we consider only the circles and points in the $(15_6 10_9)$ configuration of spheres and points, we obtain a $(15_4 10_6)$ configuration of circles and points with symmetry group S_6. The three-dimensional nature of the configuration was an essential factor in its construction in section 6; we shall now consider whether this $(15_4 10_6)$ configuration exists in a plane. The circles are

ABDE BCHJ AFHK
ACDF DEGH BDJK
(9) *BCEF DFGJ BFGK*
ABGH EFHJ CDHK
ACGJ AEJK CEGK.

If we omit the point K and the six circles through K, we obtain a (9_4) configuration that can be expressed symbolically by the square array

$$\begin{matrix} A & B & C \\ D & E & F \\ G & H & J \end{matrix} \tag{10}$$

Any four points forming the vertices of a horizontal-vertical rectangle in the array lie on a circle of the (9_4) configuration.

THEOREM 1 [5, Theorems 3, 4]. *Given such a* (9_4) *configuration, the circles ABC, DEF, GHJ are coaxal, and the circles ADG, BEH, CFJ belong to the orthogonal coaxal system; conversely, given three circles of a coaxal system and three circles of the orthogonal system, if we choose nine of their eighteen points of intersection in a suitable manner (e.g. as in Fig.* 3 *where a limiting point of one system has been taken to be at infinity) they will lie by fours on nine circles to form a* (9_4) *configuration.*

THEOREM 2 [5, §4]. *If nine points lie by fours on eight circles in the manner of a* (9_4) *configuration, then the ninth circle of the configuration exists also.*

The author has shown [5, Theorem 5] that, in the (9_4) configuration given by (10), the six circles *AEJ, AFH, BDJ, BFG, CDH, CEG* (each circle containing one point from each row and column of the array) have a common radical centre, P say. This means that P has the same power with respect to the six circles; if this power is zero, then P lies on all the circles, so we can take $P = K$ to obtain the $(15_4 10_6)$ configuration. This shows that if a (9_4) configuration satisfies a single extra condition we can obtain a $(15_4 10_6)$ configuration, but it does not tell us how to construct such a (9_4) configuration, so we use a different approach.

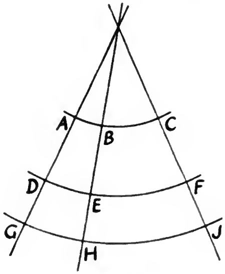

Figure 3

Supposing that the $(15_4 10_6)$ configuration exists, let us omit A and the six circles through A; we then obtain a (9_4) configuration

$$\begin{matrix} B & K & C \\ F & G & E \\ J & D & H \end{matrix} \tag{11}$$

whence FGE is orthogonal to FBJ by Theorem 1. We shall show that (9_4) configurations (10) exist with this extra orthogonality property, and that this property is sufficient to ensure the existence of K.

In Fig. 4 the lines DG, EH, FJ meet at O, and DEF, GHJ are circles with centre O. The unique circle through F and J orthogonal to FGE meets HE twice, at B and B' say. (If F and J lie on the same side of HE, this orthogonal circle may not meet HE.) The circle through B with centre O meets DG, FJ at A, C, say. We now have three circles in each of two orthogonal coaxal systems (the limiting points of one system being O and the point at infinity), so we have a (9_4) configuration of type (10) with FGE orthogonal to FBJ.

Now $OB.OB' = OF.OJ = OE.OH = OD.OG = k$ say. Hence inversion in the circle with centre O and radius $\sqrt{k}$ (an imaginary circle in Fig. 4) maps the circle $FBJB'$ to itself and FGE to JDH. But inversion preserves orthogonality, so JDH is orthogonal to FBJ.

The reflection in the internal bisector of angle EOF maps FGE and JDH to themselves and FBJ to ECH. Hence FBJ, ECH are both orthogonal to FGE, JDH, so we have two circles from each of two orthogonal coaxal systems.

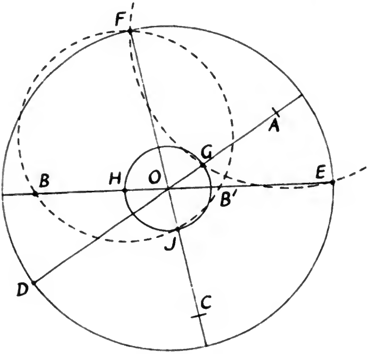

Figure 4

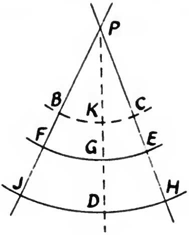

Figure 5

Assume without loss of generality that *FBJ*, *ECH* meet twice, at *P* and *Q*, and invert *Q* to infinity; we then have Fig. 5, in which *FGE*, *JDH* are circles with centre *P*. Since *BCEF* are concyclic, *B* and *C* lie on a circle with centre *P*; since *FGJD* are concyclic, the line *GD* passes through *P*. Let this circle and line meet at *K* as shown; then we have a (9_4) configuration of type (11), giving four circles through *K*, namely *BFGK*, *CEGK*, *BDJK*, *CDHK*.

To show that the (9_4) configuration

$$\begin{array}{ccc} A & H & C \\ F & K & D \\ G & B & J \end{array}$$

exists, we observe that eight circles of this configuration already are known to exist, so by Theorem 2 the ninth circle exists; i.e. *AFHK* are concyclic. Similarly by considering

$$\begin{array}{ccc} A & E & C \\ J & K & G \\ D & B & F \end{array}$$

we see that *AEJK* are concyclic. We now have all fifteen circles of a $(15_4 10_6)$ configuration.

The (9_4) configuration has four degrees of freedom with respect to the inversive group; the $(15_4 10_6)$ configuration has three, and its symmetry group is S_6.

8. **Ten Money–Coutts configurations.** The author's original interest in (9_4) configurations arose from the following result.

THEOREM 3 [5, Theorem 4]. *The nine circles of a* (9_4) *configuration meet in pairs in eighteen points other than the points of the configuration. These eighteen*

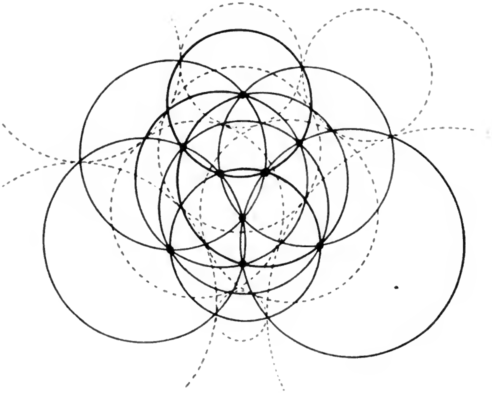

Figure 6

points lie by fours on nine circles, two of which touch at each of the eighteen points. These circles can be oriented to form a configuration of nine anti-tangent cycles, each anti-touching four others (a Money–Coutts configuration).

The situation is illustrated in Fig. 6. Now a $(15_4 10_6)$ configuration contains ten (9_4) configurations, obtained by omitting any one of the ten points and the six circles through it. Each (9_4) configuration gives a Money–Coutts configuration.

THEOREM 4. *The* 90 *circles in the ten Money–Coutts configurations derived from a* $(15_4 10_6)$ *configuration coincide in pairs to give a configuration of* 45 *circles touching by threes at* 60 *points.*

Proof. Some of the statements in this proof can be visualized by labelling the points in Fig. 6; they are proved in [5]. Consider the (9_4) configuration (10). We shall use the notation $BCEF \cap DEGH$ to denote the point of intersection, *other than E*, of the circles $BCEF$ and $DEGH$, etc.. The four points $BCEF \cap DEGH$, $DEGH \cap BCHJ$, $BCHJ \cap DFGJ$, $DFGJ \cap BCEF$ lie on a circle, which we shall denote by a. We define similarly a total of nine circles $a, b, c, \ldots, j$ forming a Money–Coutts configuration, with one circle associated with each of the nine points. Two circles touch if the corresponding points lie in distinct rows and distinct columns of the array (10); for instance, a and e touch at $BCHJ \cap DFGJ$.

Since this Money–Coutts configuration is obtained by omitting K from the $(15_4 10_6)$ configuration, we shall denote it by K and shall rename its nine circles $ka, kb, \ldots, kj$. The (9_4) configuration given by the array (11) gives the nine circles $ab, ac, \ldots, ak$ of the Money–Coutts configuration A; etc.. We easily check that $ka = ak$, etc., so there are 45 circles rather than 90.

We have seen that the circles ka and ke touch at $BCHJ \cap DFGJ$. The circles ek and ea touch at this same point, and so do ae and ak. Hence ka, ek, ae all touch at this point. We thus have a configuration of 45 circles touching by threes at 60 points, four of the points lying on each circle.

The converse of Theorem 3 is true: any Money–Coutts configuration has an associated (9_4) configuration from which it is derived. It is therefore easily seen that any configuration of 45 circles as described in Theorem 4 has an associated $(15_4 10_6)$ configuration from which it is derived. Such configurations therefore have three degrees of freedom and symmetry group S_6. We shall call them *T-configurations.* The particular T-configuration shown in Fig. 8 has six extra accidental points of triple contact, shown by black dots.

The cycles of a Money–Coutts configuration are anti-tangent; we cannot re-orient some of them to make them all tangent. However, if three circles all touch at the same point as in Theorem 4 we cannot orient them so that each anti-touches the other two. I conjecture that we can orient the cycles in each of the ten Money–Coutts configurations in such a way that the common circle of two Money–Coutts configurations is oriented in opposite directions in the two

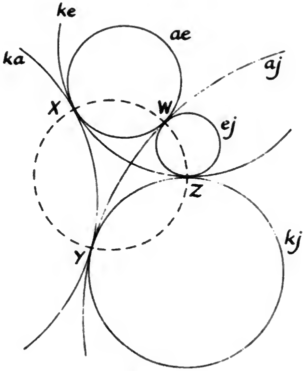

Figure 7

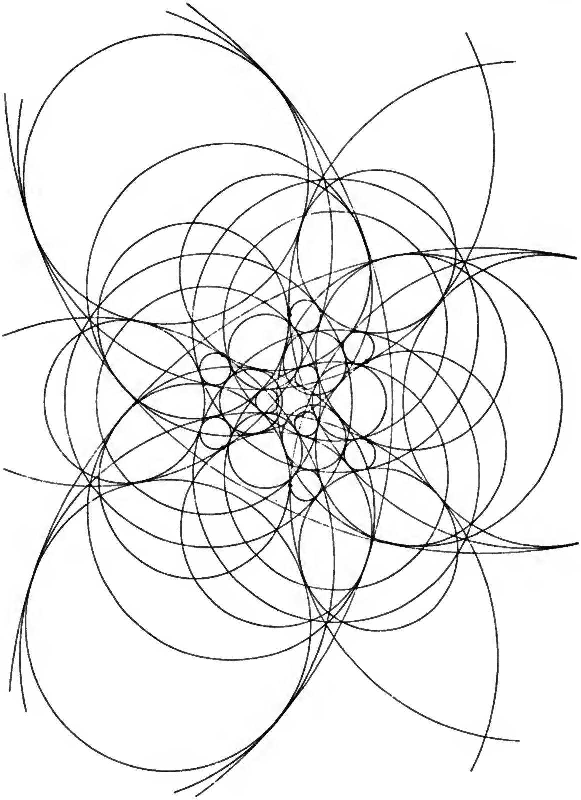

Figure 8

configurations (e.g. *ka* and *ak* are opposite cycles). We should then have a configuration of 45 bicycles.

Write $BCHJ \cap DFGJ = X$, $DEGH \cap BCEF = Y$, $BFGK \cap CDHK = W$, $ACDF \cap ABGH = Z$. The cycles *ka* and *ke* anti-touch at X, *ka* and *kj* at Y, *ke* and *kj* at Z (Fig. 7). Hence the circle XYZ is orthogonal to *ka*, *ke*, *kj*. Similarly WXY is orthogonal to *ae*, *aj*, *ak* and WXZ is orthogonal to *ea*, *ej*, *ek*. Hence $WXYZ$ are concyclic and this circle is orthogonal to *ae*, *aj*, *ak*, *ej*, *ek*, *jk* (Fig. 7). The 60 points lie by fours on fifteen such *orthocircles.*

The eighteen points of any one of the Money–Coutts configurations lie on six of these orthocircles, and these six have a common orthogonal circle [5, Theorem 1], the *base circle* of the configuration, which may be imaginary. A T-configuration contains fifteen orthocircles and ten base circles, each orthocircle orthogonal to four base circles and each base circle orthogonal to six orthocircles: a type of $(15_4 10_6)$ configuration in which the incidence relation is orthogonality.

9. **The connection with the Pascal configurations.** The circle $ABDE$ in the three-dimensional $(15_6 10_9)$ configuration is the intersection of the three spheres $\tau_{04}, \tau_{12}, \tau_{35}$, so we can denote $ABDE$ by the *syntheme* $(04, 12, 35)$ using Sylvester's terminology [1, p. 220], a syntheme being an unordered triplet of unordered pairs or *duads.* A point, such as $A = Q_{015.234}$, lies on a circle if and only if the numbers in each triplet of the suffix belong one to each duad in the corresponding syntheme.

A typical Money–Coutts circle in section 8 is *ka* or *ak*. Now K and A are $Q_{045.123}$ and $Q_{015.234}$, so we may denote *ka* by $(05, 23)$. Similarly each of the 45 circles is denoted by an unordered pair of duads. The three circles *ka*, *ke*, *ae* touch at a common point. These circles are $(05, 23)$, $(13, 45)$, $(01, 24)$; we shall denote their common point by 013245, where only the *cyclic* order of the six numbers is important, and the opposite cyclic order 542310 represents the same point. The notation is explained by the fact that 05 and 23 are "opposite adjacent pairs" in the cyclic sequence 013245, and so are the pairs 13, 45 and 01, 24. This point is also the point of intersection (other than $Q_{521.430}$) of the circles $(54, 23, 10)$ and $(42, 31, 05)$. The 60 possible cyclic orders give the 60 points of tangency of the 45 circles.

Now, if three circles touch at a common point, their centres lie on a line. Hence the centres of the 45 Money–Coutts circles lie by threes on 60 lines, forming a $(60_3 45_4)$ configuration of lines and points. From the notation just developed we immediately obtain an isomorphism between this configuration and the configuration of 60 Pascal lines associated with six points on a conic, meeting by fours at 45 points: if we denote six points of a conic by $0, 1, 2, 3, 4, 5,$ then for instance the pairs of opposite sides of the hexagon 013245 meet at the three points $(01, 24)$, $(13, 45)$, $(05, 23)$, which are collinear on the Pascal line of the hexagon.

The configuration obtained from six points on a conic is a special case of a more general configuration which we shall call a Cremona configuration [1, p. 220 et seq.]. This is derived from fifteen lines (obtained by projecting a certain configuration of fifteen lines and planes from 3-space onto a plane) denoted by duads formed from the numbers $0, 1, \ldots, 5$. Lines whose duads have no number in common meet at 45 points denoted by $(01, 23)$ etc., and these points lie by threes on 60 Pascal lines as described above, giving a $(60_3 45_4)$ configuration.

In a Cremona configuration the six points $(01, 23)$, $(01, 24)$, $(01, 25)$, $(01, 34)$, $(01, 35)$, $(01, 45)$ are collinear, on the line (01), but in a T-configuration the centres of the six circles $(01, 23)$, $(01, 24)$ etc. are not collinear. (This last statement is not proved; it has merely been observed from a figure.) Hence not all $(60_3 45_4)$ configurations of points and lines are derived from Cremona configurations.

Also in a Cremona configuration the 60 Pascal lines are concurrent by threes in 20 Steiner points; for instance, 012345, 032541, 052143 are concurrent. I conjecture from observation that in a T-configuration the *Steiner circle* through the points 012345, 032541, 052143 is orthogonal to the nine Money–Coutts circles that pass, three by three, through these points. A tedious proof of this would be possible using the Argand diagram as in [5, §3], but I have not found a synthetic proof.

The 60 Pascal lines are also concurrent by threes in 60 Kirkman points; for instance, 012345, 120534, 201453 are concurrent. Let us call the circle through the three corresponding points of a T-configuration a *Kirkman circle*. We shall say that the point *acebfd* is *conjugate* to *abcdef*. By writing the symbol for a given point in all possible ways (either retaining or reversing the cyclic order) we find that there are just three points conjugate to it; also, if B is conjugate to A then A is conjugate to B. For instance, the three points conjugate to 031524 are 012345, 120534, 201453. These points lie on a Kirkman circle, and my final conjecture is that this circle passes through 031524 also.

If this conjecture is correct, there is a natural one-one correspondence between the 60 points of a T-configuration and the 60 Kirkman circles: to each point corresponds the Kirkman circle through the point and its three conjugate points. Since certain of the points lie on orthocircles, Steiner circles etc., we might expect to discover properties of the corresponding Kirkman circles. I have as yet obtained nothing significant from this train of thought, but an interesting property of the conjectured Kirkman circles is given in the next section.

10. **A property of the conjectured Kirkman circles.** We shall assume in this final section that each Kirkman circle does pass through a fourth point, as described above; the 60 Kirkman circles and the 60 points of the T-configuration then form a (60_4) configuration.

Now the Pascal lines and Kirkman points in a Cremona configuration form a (60_3) configuration, but this splits up into six (10_3) configurations [1, p. 232], as is easily verified. Similarly our (60_4) configuration splits up into six (10_4) configurations. One of these consists of the ten points 012345, 013254, 014532, 015423, 021435, 024153, 025314, 031524, 034215, 043215, and the corresponding Kirkman circles. Let us denote this (10_4) configuration by K_1, and the other five by $K_2, \dots, K_6$.

The symmetry groups of all the configurations in this paper are transitive on points and circles, or points and spheres, and the (10_4) configurations are self-dual, as are (10_6) and (9_4). However, those symmetries of K_1 that fix one circle do not permute the points of that circle transitively: they all fix the corresponding point also. In this respect the (10_4) differs from the previous configurations; it is less symmetrical.

The $(15_4\,10_6)$ configuration from which a T-configuration is derived can be called the *auxiliary* configuration, consisting of the *auxiliary circles* and *auxiliary points* of the T-configuration. (This is preferable to the term "base points" used in [5].) Each of the 60 points of the T-configuration is the meet of two auxiliary circles; alternatively, two auxiliary circles pass through each of the 60 points.

If we consider those auxiliary circles that pass through the ten points of K_1, we find that there are just five of them, namely (02, 14, 35), (03, 15, 24), (04, 13, 25), (05, 12, 34), (01, 23, 45), and if we now adjoin these circles to K_1 we obtain a $(15_4\,10_6)$ configuration, A_1 say, easily seen to be isomorphic to the auxiliary configuration. Similarly from $K_2, \dots, K_6$ we obtain $(15_4\,10_6)$ configurations $A_2, \dots, A_6$. Thus our original auxiliary configuration, A say, gives rise to six configurations $A_1, \dots, A_6$. These in turn can be regarded as auxiliary configurations, each giving rise to six $(15_4\,10_6)$ configurations, one of which is found to be A in each case.

The final question is this: does this process "close up" in some manner, or does it lead to an infinite number of $(15_4\,10_6)$ configurations?

Note: Since the acceptance of this paper, the author has verified the conjectures made in the paper. The proofs and further results about T-configurations will be presented in a subsequent paper.

REFERENCES

1. H. F. Baker, *Principles of Geometry, vol.* 2, Cambridge 1922.
2. H. S. M. Coxeter, *Loxodromic sequences of tangent spheres*, Aeq. Math. **1** (1968), 104–121.
3. —— *Introduction to geometry, 2nd ed.*, New York 1969.
4. —— Problem 500, Crux Mathematicorum **5** (1979) 293.
5. J. F. Rigby, *On the Money–Coutts configuration of nine anti-tangent cycles*, Proc. London Math. Soc. (3) **43** (1981) 110–132.
6. Asia Weiss, *On Coxeter's loxodromic sequences*, in *The geometric vein*, Springer, N.Y. (to appear).

UNIVERSITY COLLEGE
CARDIFF CF1 1XL, WALES

Canad. Math. Bull. Vol. **25** (2), 1982

INTEGRAL REPRESENTATION BY BOUNDARY VECTOR MEASURES

BY

PAULETTE SAAB

ABSTRACT. In this paper we show that if X is a compact Hausdorff space, A is an arbitrary linear subspace of $C(X, C)$, and if E is a Banach space, then each element L of $(A\otimes E)^*$ can be represented by a boundary E^*-valued vector measure of the same norm as L.

Introduction. All the results obtained in this paper are valid for real and complex Banach spaces. However we shall deal only with complex Banach spaces.

Let X be a compact Hausdorff space and let E be a Banach space with dual E^*. Let $C(X, E)$ denote the Banach space of all continuous E-valued functions defined on X under the supremum norm. In [6] O. Hustad showed that if A is a linear subspace of $C(X, C)$ that separates the points of X and contains the constant functions, then each continuous linear function l on A can be represented by a "boundary measure" that has the same norm as l. Later Choquet [2] and Fuhr and Phelps [5] independently extended Hustad's theorem to the case in which the subspace A does not contain the constant functions. In [9], we showed that if the compact space X is metrizable and E is an arbitrary Banach space, then it is possible to extend Hustad's theorem to those subspaces of $C(X, E)$ that are of the form $A\otimes E$, where A is a linear subspace of $C(X; C)$ that separates points of X and $A\otimes E$ is the closed linear subspace of $C(X, E)$ generated by elements of the form $a\otimes v$, with a in A and v in E and where for all x in X we have;

$$a\otimes v(x)=a(x)\cdot v$$

In this paper we shall prove, using the technique we developed in [8], an extension of our result in [9]. Namely, we shall show that for any compact Hausdorff space X and any linear subspace A of $C(X, C)$, each continuous linear functional L on $A\otimes E$ can be represented by a boundary E^*-valued vector measure that has the same norm as L.

First let us collect some notations.

Received by the editors March 17, 1980.

A.M.S. (1970) subject classification: Primary 46A55, 46E40.

If V is a Banach space, we shall denote by V^* its topological dual.

It is known [7] that if Y is a compact Hausdorff space and E is a Banach space, the dual of $C(Y, E)$ is isometrically isomorphic to $M(Y, E^*)$, the space of w^*-regular E^*-valued vector measures defined on the σ-field of Borel subsets of Y and that are of bounded variation [3].

The space of all complex valued regular Borel measures on Y will simply be denoted by $M(Y)$. The subset of $M(Y)$ consisting of probability measures (resp., the positive measures) will be denoted $M_1^+(Y)$ (resp., $M^+(Y)$).

If μ is in $M(Y, E^*)$, and x is in E, we denote by $\langle x, \mu\rangle$ the element of $M(Y)$ defined as follows:

$$\langle x, \mu\rangle(B) = \mu(B)x \quad \text{for each Borel subset } B \text{ of } X.$$

1. **A representation theorem for point separating subspaces of** $C(X, C)$. Throughout this section X is a compact Hausdorff space, A is a linear subspace of $C(X, C)$ that *separates points* of X, and E is a Banach space. We shall denote by $A \otimes E$ the closed linear subspace of $C(X, E)$ generated by elements of the form $a \otimes x$, where a is in A and x is in E. Also we shall denote by U the unit ball of A^*, by $\phi: X \rightarrow U$ the canonical map, and by T the unit circle in the plane.

DEFINITION 1.1. A measure μ in $M(X, E^*)$ is called a *boundary vector measure* for A if its variation $|\mu|$ (when carried via ϕ) is maximal for the Choquet ordering on $M^+(U)$ [1].

The following Lemma can easily be obtained using the characterization of maximal measures on compact convex sets [1, 27.4].

LEMMA 1.2. *A vector measure μ in $M(X, E^*)$ is a boundary vector measure for A if and only if for each x in E the scalar measure $\langle x, \mu\rangle$ is a boundary measure for A.*

We are now ready to prove the main result of this section.

THEOREM 1.3. *Let X be a compact Hausdorff space, let A be a linear subspace of $C(X, C)$ that separates points of X, and let E be a Banach space. Then for each L in $(A \otimes E)^*$ there exists a vector measure μ in $M(X, E^*)$ such that*

(i) $\|\mu\| = \|L\|$,

(ii) $\int_X b\, d\mu = L(b)$ *for all b in $A \otimes E$, and*

(iii) *the measure μ is a boundary vector measure for A.*

Proof. If $A(U, E)$ denotes the Banach space of all continuous affine E-valued functions on U, then $A \otimes E$ embeds isometrically in $A(U, E)$ as follows: For $b=\sum_{i=1}^{n} a_i\otimes x_i$ we let $j(b)$ denote the element of $A(U, E)$ defined by

$$j(b)(a^*)=\sum_{i=1}^{n} a^*(a_i)\cdot x_i, \quad \text{for all } a^* \text{ in } U.$$

The mapping j is obviously linear. It is an isometry since the set of extreme points of U is included in $T\cdot\phi(X)$. Let L be an element of $(A\otimes E)^*$. With the help of the Hahn–Banach theorem, pick Φ in $A(U, E)^*$ such that, Φ when restricted to $j(A\otimes E)$, is equal to L and $\|\phi\|=\|L\|$.

By [8] the functional L can be represented by a measure λ in $M(U, E^*)$ such that

(i) $L(b)=\int_U j(b)\, d\lambda$ for all b in $A\otimes E$,

(ii) $\|\lambda\|=\|L\|$, and

(iii) the variation $|\lambda|$ of λ is maximal for the Choquet ordering on $M^+(U)$.

Since the measure $|\lambda|$ is maximal it is supported by $T\cdot\phi(X)$ [5]. Let $s: T\cdot\phi(X)\rightarrow T\times X$ be the Borel selection map defined by Fuhr and Phelps [5, Lemma 7.2]. Denote by $s(\lambda)$ the E^*-valued set function defined on Borel subsets of $T\times X$ as follows:

$$s(\lambda)(B)=\lambda(s^{-1}(B)) \quad \text{for each Borel subset } B \text{ of } T\times X.$$

It is easily checked that $s(\lambda)$ is in $M(T\times X, E^*)$. Let μ be equal to $H_*s(\lambda)$, where for every ν in $M(T\times X, E^*)$ and every f in $C(X, E)$

$$H_*\nu(f)=\int_{T\times X} t\cdot f\, d\nu.$$

It can easily be checked that μ is in $M(X, E^*)$, and that for each x in E

$$\langle x, H_*s(\lambda)\rangle=Hs(\langle x, \lambda\rangle)$$

where H is Hustad's map see [6] or [5].

We claim that μ is our required element. For this, note that for each a in A and for each e in E

$$\mu(a\otimes e)=\langle e, H_*s(\lambda)\rangle(a)=Hs(\langle e, \lambda\rangle)(a)$$
$$=\int_{T\cdot\phi(X)} j(a\otimes e)\, d\lambda=L(a\otimes e).$$

This shows that $H_*s(\lambda)$ and L agree on $A\otimes E$. This proves (ii). To prove (i), it is easy to check that

$$\|H_*s(\lambda)\|\le\|\lambda\|=\|L\|,$$

hence $\|H_*s(\lambda)\|=\|L\|$.

Finally, since $|\lambda|$ is maximal for the Choquet ordering on $M^+(U)$, then for each x in E, the scalar measure $|\langle x, \lambda\rangle|$ is also maximal. This implies that

$Hs(\langle x, \lambda\rangle)$ is a boundary measure for A. An appeal to Lemma 1.2 shows that the vector measure $H_*s(\lambda)$ is a boundary vector measure for A since for each x in E $\langle x, H_*s(\lambda)\rangle = Hs(\langle x, \lambda\rangle)$. This completes the proof.

2. **A representation theorem for arbitrary subspaces of** $C(X, C)$. We shall now proceed to prove Theorem 1.3 for an arbitrary linear subspace A of $C(X, C)$.

For each a in A denote by $\tilde{a}$ the element of $C(\phi(X), C)$ defined by:

$$\tilde{a}(\phi(x)) = a(x) \quad \text{for all } x \text{ in } X.$$

It is clear the element $\tilde{a}$ is well defined. Denote by $\tilde{A}$ the set $\{\tilde{a} : a \in A\}$. If E is a Banach space, consider $\tilde{A} \otimes E$ the corresponding linear subspace of $C(\phi(X), E)$. The spaces $A \otimes E$ and $\tilde{A} \otimes E$ are isometrically isomorphic. We can now prove the main result of this paper.

THEOREM 2.1. *Let A be an arbitrary linear subspace of $C(X, C)$ and let E be a Banach space. Then for each L in $(A \otimes E)^*$ there exists a measure μ_L in $M(X, E^*)$ such that*

(i) $\|\mu_L\| = \|L\|$,

(ii) $\int_X b\, d\mu_L = L(b)$ *for all b in $A \otimes E$, and*

(iii) *the measure μ_L is a boundary vector measure for A.*

Proof. Let L be in $(A \otimes E)^*$ with $\|L\| = 1$. By virtue of the isometry of $A \otimes E$ and $\tilde{A} \otimes E$ we may and do assume that L is in $(\tilde{A} \otimes E)^*$. Apply Theorem 1.3 for $\tilde{A}$ and $\phi(X)$ to get a measure ν in $M(\phi(X), E^*)$ such that

(i) $\|\nu\| = \|L\|$,

(ii) $\int \tilde{b}\, d\nu = L(b)$ for all b in $A \otimes E$, and

(iii) the measure $|\nu|$ is maximal for the Choquet ordering on $M^+(U)$.

Since $|\nu|$ is in $M_1^+(\phi(X))$, there is a net of positive discrete measures $(\nu_i)_{i \in I}$ such that $\nu_i = \sum_{j=1}^{n_i} \alpha_j^i \varepsilon_{\phi(y_j^i)}$ and $\|\nu_i\| = 1$ for each i in I, and such that ν_i converges to $|\nu|$ in the weak* topology of $M_1^+(\phi(X))$.

For each $i \in I$, let

$$\mu_i = \sum_{j=1}^{n_i} \alpha_j^i \varepsilon_{y_j^i},$$

and note that the measure μ_i is in the weak* compact convex set $M_1^+(X)$. Let μ be a weak* cluster point of the net $(\mu_i)_{i \in I}$ in $M_1^+(X)$. It is a straightforward computation to show that $\phi(\mu) = |\nu|$. Since ν is in $M(\phi(X), E^*)$, it follows from [4, p. 389] that there exists a mapping $g : \phi(X) \to E^*$ that is $|\nu|$-essentially bounded by one, (scalarly) weak*-Borel measurable, and such that $\nu = g \cdot |\nu|$. Let $\mu_L = g \circ \phi \cdot \mu$ be the E^*-valued set function defined on Borel subset B of X by:

$$\mu_L(B)(x) = \int_B \langle g \circ \phi(\omega), x\rangle\, d\mu(\omega) \quad \text{for each } x \text{ in } E.$$

It is clear that μ_L is in $M(X, E^*)$ and that the variation $|\mu_L| \leq \mu$. We claim that μ_L is our required element. To this end, note that for each a in A and for each x in E we have

$$\begin{aligned}\int_X a \otimes x \, d\mu_L &= \int_X \langle g \circ \phi(\omega), a(\omega) \cdot x \rangle \, d\mu(\omega) \\ &= \int_X \langle g \circ \phi(\omega), x \rangle \tilde{a}(\phi(\omega)) \, d\mu(\omega) \\ &= \int_{\phi(X)} \tilde{a} \otimes x \, d\nu = L(a \otimes x).\end{aligned}$$

This shows that $\mu_L = L$ on $A \otimes E$. Since $|\mu_L| \leq \mu$, it follows that $\|\mu_L\| = \|\mu\| = 1$. Hence $|\mu_L| = \mu$. Finally, the measure μ_L is a boundary vector measure for A since $\phi(|\mu_L|) = |\nu|$ is maximal for the Choquet ordering on $M^+(U)$. This completes the proof.

ACKNOWLEDGEMENT. The author is grateful to Professor J. J. Uhl, Jr. for valuable discussions concerning this paper.

BIBLIOGRAPHY

1. G. Choquet, Lecture on Analysis, Volume II, Benjamin, New York, 1969. (MR40 #3253).

2. G. Choquet, Frontière-module et représentation intégrale [Résumé], Séminaire Choquet, Univ. de Paris (1971/1973), No. 8, 4 pp.

3. J. Diestel and J. J. Uhl, Jr., Vector measures. Math. Surveys, Volume 15, American Math. Soc., Providence, 1977.

4. N. Dinculeanu, Vector measures, Berlin, VEB Deutscher Verlag der Wissenschaften, 1966.

5. R. Fuhr and R. R. Phelps, Uniqueness of complex representing measures on the Choquet boundary, J. Functional Analysis 14 (1973), 1–27. (MR50 #14186).

6. O. Hustad, A norm preserving complex Choquet Theorem, Math. Scand. 29 (1971), 272–278. (MR48 #852).

7. I. Singer, Sur la meilleur approximation des fonctions abstraites continues à valeurs dans un espace de Banach. Rev. Roumaine Math. Pures Appl. 2 (1957), 245–262.

8. P. Saab, The Choquet integral representation in the affine vector-valued case. Aequationes Mathematicae 20 (1980), 252–262.

9. P. Saab, Representation intégrale dans des sous-espaces de fonctions à valeurs vectorielles. Séminaire Choquet, (1974–1975), No. 23, 10 p.

10. H. H. Schaefer, Banach Lattices and Positive Operators. Springer-Verlag, New York–Heidelberg–Berlin, 1974.

THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C., CANADA V6T 1Y4

Canad. Math. Bull. Vol. **25** (2), 1982

"TOPOLOGICALLY INDEXED FUNCTION SPACES AND ADJOINT FUNCTORS"

BY
S. B. NIEFIELD

ABSTRACT. Let *Top* denote the category of topological spaces and continuous maps. In this paper we discuss families of function spaces indexed by the elements of a topological space T, and their relationship to the characterization of right adjoints $Top/S \to Top/T$, where S is also a topological space. After reducing the problem to the case where S is a one-point space, we describe a class of right adjoints $Top \to Top/T$, and then show that every right adjoint $Top \to Top/T$ is isomorphic to one of this form. We conclude by giving necessary and sufficient conditions for a left adjoint $Top/T \to Top$ to be isomorphic to one of the form $- \times_T Y$, where Y is a space over T, and $\times_T$ denotes the fiber product with the product topology.

Introduction. A great deal has been written about "suitable" topologies on function spaces. Although most of it appeared during the past two decades, many of the problems can be traced back to a 1945 paper of R. H. Fox, entitled "On topologies for function spaces" [4]. In this paper Fox writes

> Given topological spaces X, T and Y and a function h from $X \times T$ to Y which is continuous in x for each fixed t, there is associated with h a function h^* from T to $F = Y^X$, the space whose elements are continuous functions from X to Y. The function h^* is defined as follows: $h^*(t) = h_t$, where $h_t(x) = h(x, t)$ for every x in X... It would be desirable to so topologize F that the functions h^* which are continuous are precisely those which correspond to continuous functions h. It has been known for a long time that this is possible if X satisfies certain conditions, chief among which is the condition of local compactness... several years ago, in a letter, Hurewicz proposed to me the problem of defining such a topology for F when X is not locally compact. At that time I showed by an example that this is not in general possible. Recently I discovered that, by restricting the range of T in a very reasonable way, one of the standard topologies for F has the desired property even for spaces X which are not locally compact...

The existence of a bijection between continuous h^* and continuous h is equivalent to the statement that X is *cartesian* in *Top* (i.e. the functor $- \times X$: $Top \to Top$ has a right adjoint). The characterization of cartesian spaces was dealt with by R. Brown [2], P. Wilker [13] and finally completed by B. J.

Received by the editors March 17, 1980.

Day and G. M. Kelly [3] in 1970. Day and Kelly essentially showed that a space Y is cartesian if and only if the collection $O(Y)$ of open subsets of Y is a continuous lattice in the sense of Scott [10].

The question of the necessity of local compactness of a cartesian space Y was answered affirmitively for separable metrizable spaces by Fox [4], and Hausdorff spaces by Day and Kelly [3]. For a non-Hausdorff space Y there are two notions of local compactness, the existence of a compact neighbourhood for each element of Y, and the existence of arbitrarily small compact neighborhoods for each element of Y. In [5], K. H. Hoffmann and J. D. Lawson showed that a sober space Y is cartesian if and only if Y is locally compact (in the latter sense). They also give an example of a cartesian space which is not locally compact.

Fox's idea of "restricting the range of T" led to J. Kelley's introduction of k-spaces [8], and later Steenrod's compactly generated spaces [12]. Each was an example of a *cartesian closed* subcategory of *Top* (i.e. one in which every object is cartesian). On the other hand, in "Quasi-topologies" [11], Spanier generalized the notion of topological space to obtain a cartesian closed category containing *Top*.

A third function space problem dealt with topologizing the sets $X \times Y$ and $Top(Y, X)$, for a fixed space Y, and all spaces X, so that $- \times Y$ is left adjoint to $Top(Y, -)$ as endofunctors of *Top* (where $Top(Y, X)$ denotes the set of continuous maps from Y to X). In particular, continuous maps with domain $X \times Y$ must be continuous in x for each y (c.f. the above quote of Fox). This problem was considered by R. Brown [1], [2], P. Wilker [13], and J. Isbell [6]. In "Function spaces and adjoints", Isbell showed that every adjoint pair of endofunctors of *Top* is of this form.

In [9], Day and Kelly's characterization is generalized to the category *Top/T* of spaces over a fixed space T. After hearing about these results, F. W. Lawvere mentioned Isbell's paper and suggested that the methods of [9] be used to characterize adjoint pairs of endofunctors of *Top/T*, as Isbell did for *Top*.

The author is grateful to Lawvere for this suggestion, and to Barry Mitchell for his helpful comments and encouragement.

1. **The reduction.**

Throughout this section all categories have finite limits. If $\mathbf{A}$ is a category, then $|\mathbf{A}|$ denotes the class of objects of $\mathbf{A}$, and $\mathbf{A}(X, Y)$ denotes the set of morphisms from X to Y in $\mathbf{A}$. If $F: \mathbf{A} \to \mathbf{B}$ is a functor, and G is a right adjoint for F, we write $F \dashv G$.

If $T \in |\mathbf{A}|$, then $\mathbf{A}/T$ denotes the category whose objects are $\mathbf{A}$-morphisms $X \to T$, and morphisms are commutative triangles. An object X over T will sometimes be denoted by X without explicit reference to the projection $X \to T$. Note that if 1 denotes the terminal object of $\mathbf{A}$, then $\mathbf{A}/1 \cong \mathbf{A}$.

If $p: S \to T$ is a morphism of $\mathbf{A}$, then composition with p induces a functor $\Sigma_p: \mathbf{A}/S \to \mathbf{A}/T$, which is left adjoint to the functor $p^*: \mathbf{A}/T \to \mathbf{A}/S$ defined by pulling back along p. If $T = 1$ and p is the unique map $S \to 1$, Σ_p and p^* are considered as functors $\mathbf{A}/S \to \mathbf{A}$ and $\mathbf{A} \to \mathbf{A}/S$, respectively and are denoted by Σ_S and S^*, respectively. Note that Σ_S is the forgetful functor.

Given a functor $F: \mathbf{B} \to \mathbf{A}/S$, by composing with Σ_S, we obtain a functor $\bar{F}: \mathbf{B} \to \mathbf{A}$ together with a morphism $\bar{F}1 \to S$ (i.e. $F1$ considered as a morphism of $\mathbf{A}$). A natural transformation $\eta: F \to G$ induces a natural transformation $\bar{\eta}: \bar{F} \to \bar{G}$ such that the diagram

$$\begin{tikzcd}
\bar{F}1 \arrow[rr, "\bar{\eta}"] \arrow[dr, "F1"'] & & \bar{G}1 \arrow[dl, "G1"] \\
& S &
\end{tikzcd}$$

commutes. Moreover, it is not difficult to show that a functor $F: \mathbf{B} \to \mathbf{A}/S$ has a right adjoint if and only if $\Sigma_S \circ F$ has a right adjoint [9, proposition 1.1].

Therefore, the category of left adjoints $\mathbf{B} \to \mathbf{A}/S$ is isomorphic to the category of left adjoints $F: \mathbf{B} \to \mathbf{A}$ equipped with a morphism $F1 \to S$, with the obvious morphism. In particular, taking $\mathbf{B} = \mathbf{A}/T$, we see that to characterize left adjoints $\mathbf{A}/T \to \mathbf{A}/S$ it suffices to determine all left adjoints $\mathbf{A}/T \to \mathbf{A}$.

2. **A class of adjoint pairs.** First, we consider a general construction. Let $G: \mathbf{A} \to \mathit{Sets}/T$ be a functor, where $\mathbf{A}$ is any category. In addition, suppose Z_0 is a weak terminal object of $\mathbf{A}$ (i.e. $\mathbf{A}(Z, Z_0)$ is non-empty, for all $Z \in |A|$), and a topology is given for GZ_0 such that the projection to T is continuous. If Z is any object of $\mathbf{A}$, let $\bar{G}Z$ denote the set GZ with the topology induced by the collection of all maps $G\alpha$ such that $\alpha: Z \to Z_0$ is a morphism of $\mathbf{A}$. Using the fact that Z_0 is a weak terminal object, we see that the projection $\bar{G}Z \to T$ is continuous. Thus, $\bar{G}: \mathbf{A} \to \mathit{Top}/T$ is a functor. Note that the identity $GZ \to \bar{G}Z$ need not be continuous.

Now, suppose that we are given a family $\{Y_t\}$ of spaces indexed by the elements of a space T. For the moment, we impose no topology on the set $Y = \coprod_{t \in T} Y_t$. If Z is a space, let $\mathit{Top}(\{Y_t\}, Z) = \coprod_{t \in T} \mathit{Top}(Y_t, Z)$, i.e. the collection of pairs (σ, t), where $\sigma: Y_t \to Z$ is a continuous map. Then it is easy to see that $\mathit{Top}(\{Y_t\}, -)$ defines a functor $\mathit{Top} \to \mathit{Sets}/T$.

Let $\mathbf{2}$ denote the *Sierpinski space* $\{0, 1\}$ with $\{1\}$ open but not $\{0\}$, and let $O(\{Y_t\}) = \coprod_{t \in T} O(Y_t)$, where $O(Y_t)$ is the collection of open subsets of Y_t. If Y is any space, there is a bijection between open subsets of Y and continuous maps $Y \to \mathbf{2}$. Thus, $\mathit{Top}(\{Y_t\}, \mathbf{2})$ can be identified with $O(\{Y_t\})$, and in accordance with the above principle with $Z_0 = \mathbf{2}$, a topology $\mathscr{K}$ on $O(\{Y_t\})$, such that the projection to T is continuous, induces a functor $\mathit{Top} \to \mathit{Top}/T$ which we shall denote by $\mathit{Top}_{\mathscr{K}}(\{Y_t\}, -)$. Explicitly, the topology on $\mathit{Top}_{\mathscr{K}}(\{Y_t\}, Z)$ has as a subbase consisting of the collection of subsets of the form

$$\langle K, W\rangle = \{(\sigma, t) \mid (\sigma^{-1}W, t) \in K\}$$

where $K \in \mathscr{K}$ and W is open in Z.

If A is a set, let $O(\{Y_t\})^A$ denote the product over T of A copies of $O(\{Y_t\})$ with the product topology. Elements of $O(\{Y_t\})$ will be denoted by $\langle\langle U_\alpha, t\rangle\rangle_{\alpha\in A}$. Then $\mathscr{K}$ is a *topological topology* if the map $\bigcup_A : O(\{Y_t\})^A \to O(\{Y_t\})$ is continuous for every set A, and the map $\bigcap_A : O(\{Y_t\})^A \to O(\{Y_t\})$ is continuous for every finite set A. If A is empty, then $\bigcup_A, \bigcap_A : T \to O(\{Y_t\})$ are given by $t \mapsto (\phi, t)$ and $t \mapsto (Y_t, t)$, respectively. Hence, $\bigcup_\phi$ and $\bigcap_\phi$ are continuous if and only if for every K open in $O(\{Y_t\})$

$$\{t \mid (\phi, t) \in K\} \quad \text{and} \quad \{t \mid (Y_t, t) \in K\}$$

are open in T, respectively. In fact, this is precisely what we need to show that $O(\{Y_t\}) \to Top_{\mathscr{K}}(\{Y_t\}, 2)$ is continuous.

If $p : X \to T$ is a space over T, let $p \times \{Y_t\} = \{(x, y) \mid y \in Y_{px}\}$. We shall say that a subset W of $p \times \{Y_t\}$ is *open* if $W[x] = \{y \mid (x, y) \in W\}$ is open in Y_{px}, and $x \mapsto W[x]$ defines a continuous map $X \to O(\{Y_t\})$. If W_α, $\alpha \in A$ is a family of open subsets of $p \times \{Y_t\}$, then $\bigcup W_\alpha$ and $\bigcap W_\alpha$ (if A is finite) correspond to the composites $X \to O(\{Y_t\})^A \xrightarrow{\bigcup_A} O(\{Y_t\})$ and $X \to O(\{Y_t\})^A \xrightarrow{\bigcap_A} O(\{Y_t\})$, respectively. Thus, if $\mathscr{K}$ is a topological topology, then the open subsets of $p \times \{Y_t\}$ define a topology, and we shall denote the resulting space by $p \times_{\mathscr{K}} \{Y_t\}$. Conversely, if the open subsets of $p \times \{Y_t\}$ form a topology for every space X over T, then taking $X = O(\{Y_t\})^A$ and the identity $X \to O(\{Y_t\})^A$, we see that $\mathscr{K}$ is a topological topology. In this case, it is easy to see that a continuous map $X \to X'$ over T induces a continuous map $p \times_{\mathscr{K}} \{Y_t\} \to p' \times_{\mathscr{K}} \{Y_t\}$, and it follows that $-\times_{\mathscr{K}} \{Y_t\}$ defines a functor $Top/T \to Top$.

Proposition 2.1. *If $\{Y_t\}$ is a family of spaces indexed by the elements of T, and $\mathscr{K}$ is a topological topology on $O(\{Y_t\})$, then $-\times_{\mathscr{K}}\{Y_t\} \dashv Top_{\mathscr{K}}(\{Y_t\}, -)$.*

Proof. Suppose that X is a space over T with projection p, and Z is a space. Let $\eta : X \to Top_{\mathscr{K}}(\{Y_t\}, p \times_{\mathscr{K}} \{Y_t\})$ and $\varepsilon : Top_{\mathscr{K}}(\{Y_t\}, Z) \times_{\mathscr{K}} \{Y_t\} \to Z$ be given by $\eta(x) = ((x, -), px)$ and $\varepsilon((\sigma, t), y) = \sigma y$.

If W is open in $p \times_{\mathscr{K}} \{Y_t\}$, let $w : X \to O(\{Y_t\})$ and $\chi_W : X \times_{\mathscr{K}} \{Y_t\} \to 2$ be the corresponding continuous maps. Consider the diagram

$$\begin{tikzcd}
X \arrow[r, "\eta"] \arrow[d, "w"'] & Top_{\mathscr{K}}(\{Y_t\}, p \times_{\mathscr{K}} \{Y_t\}) \arrow[d, "Top_{\mathscr{K}}(\{Y_t\}, \chi_W)"] \\
O(\{Y_t\}) \arrow[r, phantom, "\cong"] & Top_{\mathscr{K}}(\{Y_t\}, 2)
\end{tikzcd}$$

If $K\in\mathscr{K}$, then $\eta^{-1}(\langle K, W\rangle)=\eta^{-1}\,Top_{\mathscr{K}}(\{Y_t\},\chi_W)^{-1}(K)=w^{-1}(K)$, and hence η is continuous since w is.

Now, suppose that W is open in Z, and $\chi_W: Z\to\mathbf{2}$ is the corresponding continuous map. Then $\varepsilon^{-1}(W)=\varepsilon^{-1}\chi_W^{-1}(1)$ is open in $Top_{\mathscr{K}}(\{Y_t\},Z)\times_{\mathscr{K}}\{Y_t\}$ since the induced map $Top_{\mathscr{K}}(\{Y_t\},Z)\to O(\{Y_t\})\cong Top_{\mathscr{K}}(\{Y_t\},\mathbf{2})$ is $Top_{\mathscr{K}}(\{Y_t\},\chi_W)$. To complete the proof one checks that the adjunction identities hold relative to η and ε.

3. **The characterization.** If $p: X\to T$ is a continuous map and $t\in T$, then the *fiber of X over t* is the set $X_t=p^{-1}t$ with the subspace topology. If $t: 1\to T$ denotes the constant t valued map, then there is a bijection between elements of X_t, and morphisms $t\to p$ of Top/T.

Let $G: Top\to Top/T$ be a functor with a left adjoint F. Then there is a natural bijection

$$\theta: Top(Fp, Z)\to Top/T(p, GZ)$$

which is natural in $p: X\to T$ and Z. Taking $p=t$, we see that $(GZ)_t$ can be identified with $Top(Ft, Z)$. Applying naturality to one point embeddings $x: 1\to X$, considered as morphisms $t\to p$, we see that $\theta(f)(x)$ is identified with $Ft\xrightarrow{Fx}Fp\xrightarrow{f}X$, where $x\in X_t$.

LEMMA 3.1. *Every space can be embedded as a subspace of a product of copies of the Sierpinski space $\mathbf{2}$ and the indiscrete space 2.*

Proof. It is well known that every T_0 space can be embedded in a product of $\mathbf{2}$'s. For a non-T_0 space, copies of the indiscrete space 2 are added to "separate" points.

LEMMA 3.2 (Isbell [6]). *If F is a left adjoint endofunctor of Top, then F followed by the forgetful functor Top $\to$ Sets can be expressed in the form $-\times F1$.*

Proof. If X is a space, then since F preserves epimorphisms, we have a continuous surjection $F(\coprod_{x\in X}1)\to FX$. Then since F commutes with coproducts we have a natural surjection

$$X\times F1\to FX \tag{1}$$

It is easy to see that (1) cannot identify distinct points with the same first coordinate, since every element of X is a retract as a space.

Suppose D is a directed set, and $X=D\coprod 1$ with the following topology. A subset U of X is open if $1\notin U$, or if $1\in U$ and U contains a subset of the form $\{x\in D \mid x\geq d\}$, for some $d\in D$. For any distinct pair of elements, X can be written as a coproduct separating the given element. Then, since F preserves coproducts, it follows that if X is of this form, (1) cannot identify two elements with different first coordinates.

Finally, using the fact that directed sets (i.e. nets) are enough to determine topologies we see that every space can be expressed as a colimit of "directed sets", and since both sides of (1) are colimit preserving it follows that (1) is a bijection for any space X.

Note that Isbell's proof of the above lemma uses Hausdorff ultraspaces rather than directed sets.

Let T-*Top* denote the category whose objects are pairs $(\{Y_t\}, \mathcal{K})$, where $\{Y_t\}$ is a family of spaces indexed by the elements of T, and $\mathcal{K}$ is a topological topology on $O(\{Y_t\})$. A morphism $f:(\{Y_t\}, \mathcal{K}) \to (\{Y'_t\}, \mathcal{K}')$ is a family of continuous maps $f_t: Y'_t \to Y_t$ such that the induced map $O(\{Y_t\}) \to O(\{Y'_t\})$ is continuous.

THEOREM 3.3. *T-Top is equivalent to the category of right adjoints $Top \to Top/T$ and natural transformations.*

Proof. Let Γ denote the following functor. If $(\{Y_t\}, \mathcal{K})$ is an object of T-*Top*, by Lemma 2.1, $\Gamma(\{Y_t\}, \mathcal{K}) = Top_{\mathcal{K}}(\{Y_t\}, -)$ is a right adjoint $Top \to Top/T$. A morphism $f:(\{Y_t\}, \mathcal{K}) \to (\{Y'_t\}, \mathcal{K}')$ induces a natural transformation $\Gamma(f): Top_{\mathcal{K}}(\{Y_t\}, -) \to Top_{\mathcal{K}'}(\{Y'_t\}, -)$ given by $\Gamma(f)_Z((\sigma, t)) = (\sigma \circ f_t, t)$, where $\sigma: Y_t \to Z$ is a continuous map.

First, we claim that Γ is full and faithful. A natural transformation $\eta: Top_{\mathcal{K}}(\{Y_t\}, -) \to Top_{\mathcal{K}'}(\{Y'_t\}, -)$ induces a natural transformation $\bar{\eta}: - \times_{\mathcal{K}'} \{Y'_t\} \to - \times_{\mathcal{K}} \{Y_t\}$, and hence a family of continuous maps $\bar{\eta}_t: Y'_t \to Y_t$, since $t \times_{\mathcal{K}'} \{Y'_t\} = Y'_t$ and $t \times_{\mathcal{K}} \{Y_t\} = Y_t$. It is easy to see that the corresponding map $O(\{Y_t\}) \to O(\{Y'_t\})$ is precisely η_2. Therefore Γ is full since $\{\bar{\eta}_t\}: (\{Y_t\}, \mathcal{K}) \to (\{Y'_t\}, \mathcal{K}')$ is a morphism, and $\Gamma(\{\bar{\eta}_t\}) = \eta$. To see that Γ is faithful, we note that if $f:(\{Y_t\}, \mathcal{K}) \to (\{Y'_t\}, \mathcal{K}')$, then $\Gamma(f)_{Y_t}((1_{Y_t}, t)) = (f_t, t)$.

To see that Γ is an equivalence of categories it remains to show that every right adjoint $G: Top \to Top/T$ is isomorphic to one of the form $Top_{\mathcal{K}}(\{Y_t\}, -)$, where $\mathcal{K}$ is a topological topology.

Suppose $F \dashv G$. As before, we can identify $(GZ)_t$ with $Top(Ft, Z)$. Letting $Y_t = Ft$ and identifying $O(Y_t)$ with $(G\mathbf{2})_t$ we obtain a family $\{Y_t\}$ of spaces together with a topology $\mathcal{K}$ on $O(\{Y_t\})$ such that the projection to T is continuous. If $f: Z \to \mathbf{2}$ is a continuous map, consider the diagram

(2)
$$\begin{tikzcd}
GZ \arrow[r, "\phi_Z"] \arrow[d, "Gf"'] & Top_{\mathcal{K}}(\{Y_t\}, Z) \arrow[d, "Top_{\mathcal{K}}(\{Y_t\}{,}f)"] \\
G\mathbf{2} & Top_{\mathcal{K}}(\{Y_t\}, 2) \arrow[l, "\phi_2^{-1}"]
\end{tikzcd}$$

where the rows are bijections and all maps, except possibly ϕ_Z, are continuous. But $Top_{\mathcal{K}}(\{Y_t\}, Z)$ has the topology induced by all maps

$Top_{\mathscr{K}}(\{Y_t\}, f): Top(\{Y_t\}, Z) \to Top(\{Y_t\}, \mathbf{2}) \to G\mathbf{2}$, and so ϕ_Z is continuous since Gf is. In particular, $\phi_{\mathbf{2}}$ is a homeomorphism.

We would like to show that ϕ_Z is a homeomorphism for every space Z. First, we show that K is a topological topology. Consider $\mathbf{2}$ as a poset with the order topology. Then $\sup: \mathbf{2}^A \to \mathbf{2}$ is continuous for every set A, and $\inf: \mathbf{2}^A \to \mathbf{2}$ is continuous if A is finite. Applying the fact that G preserves products we get continuous maps $(G\mathbf{2})^A \xrightarrow{\sim} G(\mathbf{2}^A) \xrightarrow{G(\sup)} G\mathbf{2}$ and $(G\mathbf{2})^A \xrightarrow{\sim} G(\mathbf{2}^A) \xrightarrow{G(\inf)} G(\mathbf{2})$, and by commutativity of (2), we see 'that these composites are $\bigcup_A$ and $\bigcap_A$, respectively.

Finally, since G and $Top_{\mathscr{K}}(\{Y_t\}, -)$ preserve subspaces and products (being right adjoints), by Lemma 3.1, it suffices to show that ϕ_2 is a homeomorphism, where 2 denotes the indiscrete two point space.

Recall that $Top_{\mathscr{K}}(\{Y_t\}, Z)$ has a basic family of open subsets of the form $\langle K, W\rangle = \{(\sigma, t) \mid (\sigma^{-1}W, t) \in K\}$, where $K \in \mathscr{K}$ and W is open in Z. When $Z = 2$, it is easy to see that every open subset is of the form $\coprod_{t \in U} Top(Y_t, 2)$, where U is an open subset of T.

Now, consider the functor $Top \xrightarrow{G} Top/T \xrightarrow{t^*} Top$, where $t: 1 \to T$ is the constant t valued map. By composition of adjoints, $t^* \circ G$ has a left adjoint, and so using Lemma 3.2 it is not difficult to show that $(G2)_t = (t^* \circ G)(2)$ is indiscrete, i.e. one shows that every map $X \to (G2)_t$ is continuous. Thus, every open subset of $G2$ is of the form $\coprod_{t \in U} (G2)_t$, where $U \subseteq T$. To see that U is open, it suffices to show that $G2$ admits a global section. But $Top/T(1_T, G2) \cong Top(F1_T, 2)$ which is clearly nonempty. Therefore, ϕ_2 is a homeomorphism, and the proof is complete.

COROLLARY 3.4. *Every left adjoint F: Top/T → Top followed by the forgetful functor Top → Sets is of the form $- \times_T Y$, for some space Y over T.*

Proof. By Theorem 3.3, $F \cong \times_{\mathscr{K}}\{Y_t\}$ where $\{Y_t\}$ is a family of spaces indexed by the elements of T, and $\mathscr{K}$ is a topological topology on $O(\{Y_t\})$. Taking $Y = \coprod_{t \in T} Y_t$ with the coproduct topology and the obvious projection to T gives the desired result.

If $(\{Y_t\}, \mathscr{K})$ is an object of T-*Top* we can also obtain a topology on the set $Y = \coprod_{t \in T} Y_t$ by identifying it with $1_T \times_{\mathscr{K}} \{Y_t\}$. First, we discuss several properties of topological topologies.

Let $\{Y_t\}$ be any family of spaces indexed by the elements of T. A subset K of $O(\{Y_t\})$ is *saturated* if whenever $(U, t) \in K$ and $U \subseteq V \in O(\{V_t\})$, then $(V, t) \in K$. K has the *finite union property* (fup) if whenever $(\bigcup_A U_\alpha, t) \in K$ and $\{U_\alpha\}_{\alpha \in A} \subseteq O(Y_t)$, then $(\bigcup_F U_\alpha, t) \in K$, for some finite $F \subseteq A$. A topology $\mathscr{K}$ on $O(\{Y_t\})$ is *saturated* or has the *finite union property* if every $K \in \mathscr{K}$ does. Note that when $T = 1$, then the collection of saturated subsets of $O(Y)$ with *fup* is known as the *Scott-topology* on the lattice $O(Y)$.

Lemma 3.5. *Let $\mathcal{K}$ be a topology on $O(\{Y_t\})$ such that the projection to T is continuous. If $\bigcup_A : O(\{Y_t\})^A \to O(\{Y_t\})$ is continuous for all A (in particular, if $\mathcal{K}$ is topological), then $\mathcal{K}$ is saturated and has the finite union property.*

Proof. Suppose that $(U, t) \in K \in \mathcal{K}$, and A is an infinite set. Let $U_\alpha = U$ for every $\alpha \in A$. Then since $\bigcup_A$ is continuous and K is open in $O(\{Y_t\})$, we can find a basic open neighbourhood of $\langle(U_\alpha, t)\rangle_{\alpha\in A}$ which maps into K. If $U \subseteq V \in O(Y_t)$, then some element $\langle(V_\alpha, t)\rangle_{\alpha\in A}$ of this neighbourhood is such that $V_\alpha = U$ for all but one α, and $V_\alpha = V$ for that α. Thus, $(V_t, t) = \bigcup_A \langle(V_\alpha, t)\rangle_{\alpha\in A} \in K$, and so K is saturated.

To see that K has *fup*, suppose that $\{U_\alpha\}_{\alpha\in A} \subseteq O(Y_t)$ and $(\bigcup_A U_\alpha, t) \in K$. Then there is a basic open neighbourhood of $\langle(U_\alpha, t)\rangle_{\alpha\in A}$ which maps into K under $\bigcup_A$, and in this neighbourhood there is an element of the form $\langle(V_\alpha, t)\rangle_{\alpha\in A}$, where $V_\alpha = \varnothing$ for all but finitely many α, and $V_\alpha = U_\alpha$ otherwise. Therefore, $(\bigcup_F U_\alpha, t) \in K$, for some finite $F \subseteq A$, proving that K has *fup*.

Suppose that $(\{Y_t\}, \mathcal{K})$ is an object of T-*Top*, and let $Y = 1_T \times_{\mathcal{K}} \{Y_t\}$. It is easy to see that the projections $p \times_{\mathcal{K}} \{Y_t\} \to X$ and $p \times_{\mathcal{K}} \{Y_t\} \to Y$ are continuous for every $p : X \to T$ in *Top*/T. In particular, when $p = 1_T$ we have continuity of the projection $\pi : Y \to T$. Therefore, it follows that the identity map $p \times_{\mathcal{K}} \{Y_t\} \to X \times_T Y$ is continuous. Now, the space Y_t is not necessarily a subspace of Y, but the inclusion $Y_t \to Y$ is, in general, continuous. If $- \times_{\mathcal{K}} \{Y_t\}$ preserves embeddings, then clearly, Y_t is a subspace of Y. If points of T are *locally closed* (i.e. for every $t \in T$ there exist U open and F closed in T such that $\{t\} = U \cap F$), then an open subset U_t of Y_t comes from an open subset of Y, namely $U = U_t \cup \pi^{-1}(Y \backslash \{\bar{t}\})$ which is open in Y since the corresponding map $T \to O(\{Y_t\})$ is continuous by Lemma 3.5 and the fact that $\{t \mid (Y_t, t) \in K\}$ and $\{t \mid (\phi, t) \in K\}$ are open in T whenever $K \in \mathcal{K}$. Thus, we have the following corollary.

Corollary 3.6. *If points of T are locally closed, and F: Top/$T \to$ Top is a left adjoint, then F followed by the forgetful functor is of the form $- \times_T F(1_T)$, for some continuous map $F(1_T) \to T$. Moreover, if S is any space, then every left adjoint F: Top/$T \to$ Top/S can be expressed in the form*

$$\mathit{Top}/T \xrightarrow{\text{“}f^*\text{”}} \mathit{Top}/\Sigma_S \circ F(1_T) \xrightarrow{\Sigma_{F(1_T)}} \mathit{Top}/S$$

where $f : \Sigma_S \circ F(1_T) \to T$ is a continuous map and the underlying set of “f^*” is given by pulling back along f.

Lemma 3.7 (Isbell [7]). *If Y is a space and $\mathcal{K}$ is a T_0 topological topology on $O(Y)$, then $\{\bar{U}\} = \{V \mid V \subseteq U\}$, for every $U \in O(Y)$.*

Proof. First, we note that $\{V \mid V \subseteq U\} \subseteq \{\bar{U}\}$ since $\mathcal{K}$ is saturated. On the other hand, if $V \in \{\bar{U}\}$, then $U \in \{\overline{U \cup V}\}$ since $\mathcal{K}$ is saturated, and $U \cup V \in \{\bar{U}\}$

since $\bigcup : O(Y)\times O(Y)\to O(Y)$ is continuous. But $\mathcal{K}$ is T_0 topology, and so $U\cup V=U$. Therefore, $V\subseteq U$.

Suppose that Y and Z are spaces. Recall that the *pointwise topology* on $Top(Y,Z)$ is the topology induced by the collection of all evaluation maps $ev_y : Top(Y,Z)\to Z$, where $y\in Y$. In particular, the pointwise topology on $O(Y)$ is generated by all subsets of the form $\{U\in O(Y)\mid y\in U\}$, for some fixed $y\in Y$. It is not difficult to show that the pointwise topology is necessarily T_0.

COROLLARY 3.8. *Let F: $Top/T\to Top$ be a functor with a right adjoint G. Then $F\cong -\times_T Y$ for some cartesian space Y over T if and only if F preserves subspaces and $(G2)_t$ contains the pointwise topology for all $t\in T$.*

Proof. Suppose that $F\cong -\times_T Y$, for some cartesian space Y over T. Then clearly F preserves subspaces. To see that $(G2)_t$ contains the pointwise topology we note that the open subsets of $(G2)_t$ are precisely the saturated subsets of $O(Y_t)$ which have the finite union property [9].

For the converse, suppose that $(\{Y_t\},\mathcal{K})$ corresponds to the right adjoint G, and let $Y=1_T\times_{\mathcal{K}}\{Y_t\}$. Then as before, the obvious projection $\pi : Y\to T$ is continuous. Now, by Lemma 3.1, every space over T can be embedded as a subspace of one of the form $\pi_3 : 2^A\times X\times T\to T$ where A is a set, and X is an indiscrete space. Thus, since F and $-\times_T Y$ preserve subspaces, it suffices to show that the induced continuous bijection $\pi_3\times_{\mathcal{K}}\{Y_t\}\to(2^A\times X\times T)\times_T Y$ is a homeomorphism.

Suppose that W is open in $\pi_3\times_{\mathcal{K}}\{Y_t\}$ and let $w : 2^A\times X\times T\to O(\{Y_t\})$ be the continuous map given by $w(\langle s_\alpha\rangle, x, t)=\{y\mid((\langle s_\alpha\rangle, x, t), y)\in W\}$. First, we note that if $\langle s_\alpha\rangle, \langle s'_\alpha\rangle\in 2^A$ where $s_\alpha\le s'_\alpha$ for all α, and $x, x'\in X$, then by Lemma 3.7 $w(\langle s_\alpha\rangle, x, t)\subseteq w(\langle s'_\alpha\rangle, x', t)$, for every $t\in T$. Fix $((\langle s_\alpha\rangle, x, t), y)\in W$, and let K be an open subset of $O(\{Y_t\})$ such that $K_t=\{(U,t)\mid y\in U\}$. Then by continuity of w there exists $\langle m_\alpha\rangle\in 2^A$ such that $m_\alpha\le s_\alpha$, for all α, $m_\alpha=0$ for all but finitely many α, and $w(\langle m_\alpha\rangle, x, t)\in K$. Now, the map f: $Y\to\pi_3\times_{\mathcal{K}}\{Y_t\}$ defined $f(y)=((\langle m_\alpha\rangle, x, \pi y), y)$ is continuous since it arises by applying $-\times_{\mathcal{K}}\{Y_t\}$ to the obvious continuous map $T\to 2^A\times X\times T$ over T. We claim that the neighbourhood $V=(\{\langle s'_\alpha\rangle\mid m_\alpha\le s'_\alpha\}\times X\times T)\times_T f^{-1}W$ of $((\langle s_\alpha\rangle, x, t), y)$ in $(2^A\times X\times T)\times_T Y$ is contained in W. To see this note that if $((\langle s'_\alpha\rangle, x', t'), y')\in V$ then $y'\in w(\langle m_\alpha\rangle, x, t')\subseteq w(\langle s'_\alpha\rangle, x', t')$, and so $((\langle s'_\alpha\rangle, x', t'), y')\in W$.

REFERENCES

1. R. Brown, *Ten topologies on X × Y*, Quart. J. Math, Oxford Series, (2) **14** (1963), 303–319.
2. R. Brown, *Function spaces and product topologies*, Quart J. Math, Oxford Series (2) **15** (1964), 228–250.
3. B. J. Day and G. M. Kelly, *On topological quotient maps preserved by pullbacks or products*, Proc. Cambridge Plulos Soc **67** (1970) 553–558.
4. R. H. Fox, *On topologies for function spaces* Bull. Amer. Math Soc. **51** (1945), 429–432.

5. K. H. Hoffman and J. D. Lawson, *The spectral theory of distributive continuous lattices*, Trans. Amer. Math. Soc 246 (1978), 285–310.

6. J. R. Isbell, *Function spaces and adjoints*, Math Scand **36**, (1975) 317–339.

7. J. R. Isbell *Meet-continuous lattices*, Sympos. Math **16**, (1975) 41–54.

8. J. L. Kelley, *General Topology*, Graduate Texts in Mathematics no. 5, Springer-Verlag, 1971.

9. S. B. Niefield, *Cartesianness: topological spaces, uniform spaces and affine schemes*, accepted by J. Pure and Applied Alg.

10. D. S. Scott, *Continuous lattices*, Springer Lecture Notes in Math **274** (1972), 97–136.

11. E. Spanier, *Quasi-topologies*, Duke Math J., **30** (1963), 1–14.

12. N. E. Steenrod, *A convenient category of topological spaces*, Michigan Math J. **14** (1967), 133–152.

13. P. Wilker *Adjoint product and hom functors in general topology*, Pacific J. Math **34** (1970), 269–283.

DALHOUSIE UNIVERSITY
HALIFAX, N.S.

Canad. Math. Bull. Vol. **25** (2), 1982

GIRTH AND INDEPENDENCE RATIO

BY

GLENN HOPKINS AND WILLIAM STATON

ABSTRACT. Lower bounds are given for the independence ratio in graphs satisfying certain girth and maximum degree requirements. In particular, the independence ratio of a graph with maximum degree Δ and girth at least six is at least $(2\Delta-1)/(\Delta^2+2\Delta-1)$. Sharper bounds are given for cubic graphs.

Professor Erdös [3] has shown that there exist graphs with very large girth and yet very small independence ratio. Such graphs must of course have very large maximum degree, for if the maximum degree is bounded above by p, then the independence ratio is bounded below by $1/(p+1)$. As a consequence of Brooks' Theorem [2], the independence ratio is bounded below by $1/p$ if one assumes maximum degree p, no complete subgraphs on $p+1$ vertices, and $p \geq 3$. More recently, Albertson, Bollobas and Tucker [1], assuming no complete subgraphs on p vertices have improved the inequality yielded by Brooks' Theorem to a strict inequality, except in two cases, which are demonstrated. Specializing to the case $p=3$, Fajtlowicz [4] and Staton [6, 7] have determined constants larger than $\frac{1}{3}$ which serve as lower bounds for the independence ratio in cubic triangle-free graphs. Staton's constant, $\frac{5}{14}$, is shown to be best possible by an example of Fajtlowicz in which this ratio is achieved.

In [5], Fajtlowicz showed that graphs with maximum degree p containing no complete subgraphs with q vertices have independence ratio at least $2/(p+q)$. When $q=3$, this says that graphs with girth at least four have independence ratio at least $2/(p+3)$. Taking this result as our point of departure, we investigate independence ratio in graphs with fixed lower bounds on girth and fixed upper bounds on the maximum degree. We concentrate primarily on cubic graphs.

We will employ the ideas and notation introduced by Fajtlowicz in [4] and extended by him in [5]. In particular, if G is a finite graph, then α_0 will be the size of a maximum independent subset of G, n will be the number of vertices of G, and the ratio α_0/n will be called the independence ratio of G. If I is a maximum independent vertex set in G, and if p is the maximum degree of G, then for $1 \leq i \leq p$, $G_i(I) = G_i$ will be the set of all vertices of G which are

---

Received by the editors March 17, 1980 and in revised form, July 15, 1980.

AMS (1980) Classification number: 05.

Key Words: Girth, independent sets, cubic.

adjacent to exactly i vertices of I. The cardinality of G_i will be denoted $\alpha_i = \alpha_i(I)$. Note that the G_i's are disjoint, so that $\sum_{i=1}^{p} \alpha_i = n - \alpha_0$. For $1 \le i \le p$, Γ_i will be the set of all vertices in I adjacent to at least one vertex of G_i. Note that the Γ_i's need not be disjoint. If X and Y are disjoint sets of vertices, then $[X, Y]$ will denote the collection of edges with one end in X and one end in Y.

If G is cubic, then we have $\alpha_1 + \alpha_2 + \alpha_3 = n - \alpha_0$, and counting the number of edges in $[I, G-I]$, we get $\alpha_1 + 2\alpha_2 + 3\alpha_3 = 3\alpha_0$. Eliminating α_2 from these two equations, and solving for the independence ratio, we get

$$\frac{\alpha_0}{n} = \frac{2}{5} - \frac{\alpha_1 - \alpha_3}{5n},$$

an equation to which we will refer frequently. In [4], Fajtlowicz shows that if G contains no triangle, and if I is a maximum independent set, then each vertex in Γ_1 has exactly one neighbor in G_1, so that $|\Gamma_1| = \alpha_1$. This observation is easily proven, and we feel free to use it in what follows.

LEMMA 1. *If G has maximum degree p, then*

$$\frac{\alpha_0}{n} \ge \frac{2}{p+2} - \frac{\alpha_1}{n(p+2)}.$$

Proof. Counting vertices, we get $\alpha_1 + \alpha_2 + \cdots + \alpha_p = n - \alpha_0$, and counting those edges with an end in I, we get $\alpha_1 + 2\alpha_2 + \cdots + p\alpha_p \le p\alpha_0$. Doubling the equation and subtracting from the inequality yields

$$(p+2)\alpha_0 - 2n \ge -\alpha_1 + \alpha_3 + 2\alpha_4 + \cdots + (p-2)\alpha_p.$$

It follows that

$$\frac{\alpha_0}{n} \ge \frac{2}{p+2} - \frac{\alpha_1}{n(p+2)} + \frac{\sum_{k=2}^{p} (k-2)\alpha_k}{n(p+2)},$$

the last summand of which is nonnegative. ■

LEMMA 2. *If G has maximum degree p, no 3-cycle, and no 5-cycle, then, for $k \ge 2$, each vertex in G_k has at least one neighbor in $I - \Gamma_1$.*

Proof. Suppose that v is a vertex in G_k and that $x_1, x_2, \ldots, x_k$ are the neighbors of v which lie in I. If each x_i is in Γ_1, then there are vertices $y_1, y_2, \ldots, y_k$ in G_1 such that x_i is adjacent to y_i for $1 \le i \le k$. Since G has no 3-cycle, no y_i is adjacent to v, and, since G has no 5-cycle, no two of the y_i's are adjacent. Since the only vertices in I adjacent to the y_i's are the x_i's, the set

$$(I - \{x_1, x_2, \ldots, x_k\}) \cup \{v, y_1, y_2, \ldots, y_k\}$$

is an independent set larger than I. This contradiction shows that at least one of the neighbors of v in I must fail to be in Γ_1. ■

THEOREM 3. *If G has maximum degree p, no 3-cycle, and no 5-cycle, then*

$$\frac{\alpha_0}{n} \geq \frac{2p-1}{p^2+2p-1}.$$

Proof. By Lemma 2, each vertex in $G_2 \cup G_3 \cup \cdots \cup G_p$ has at least one neighbor in $I-\Gamma_1$. Since each vertex in $I-\Gamma_1$ has degree at most p, it follows that

$$|I-\Gamma_1| \geq \frac{\alpha_2+\alpha_3+\cdots+\alpha_p}{p}.$$

But $|I-\Gamma_1| = \alpha_0 - \alpha_1$, and $\alpha_2+\alpha_3+\cdots+\alpha_p = n-\alpha_0-\alpha_1$. Thus

$$\alpha_0-\alpha_1 \geq \frac{n-\alpha_0-\alpha_1}{p}, \quad \text{or} \quad \alpha_1 \leq \frac{p+1}{p-1}\alpha_0 - \frac{n}{p-1}.$$

Invoking Lemma 1, we get

$$\frac{\alpha_0}{n} \geq \frac{2}{p+2} - \frac{\dfrac{p+1}{p-1}\alpha_0 - \dfrac{n}{p-1}}{n(p+2)},$$

which implies

$$\frac{\alpha_0}{n} \geq \frac{2p-1}{p^2+2p-1}. \quad \blacksquare$$

We note that the ratio $(2p-1)/(p^2+2p-1)$ is larger than the $2/(p+3)$ obtained by Fajtlowicz for triangle free graphs. For $p=3$, Theorem 3 gives an independence ratio of at least $\frac{5}{14}$. We note that in [7], it was shown that the $\frac{5}{14}$ ratio may be gotten even under the milder assumption of no triangles. However, specializing to cubic graphs, we now obtain a bound larger than $\frac{5}{14}$.

LEMMA 4. *If G is cubic and contains no 3-cycle, and if I is a maximum independent set with the property that $\alpha_3(I)$ is maximum, then $|\Gamma_1-\Gamma_3| \leq |\Gamma_1 \cap \Gamma_3|$.*

Proof. Let v be a vertex in $\Gamma_1-\Gamma_3$, and let x be its neighbor in G_1. We claim that x has no neighbor in G_2. For if y is a neighbor of x and $y \in G_2$, then $J=(I-\{v\}) \cup \{x\}$ is a maximum independent set. But $G_3(J)=G_3(I) \cup \{y\}$, which contradicts the maximality of $\alpha_3(I)$. Thus x has no neighbor in G_2, so it has two neighbors in G_1, say y and z. Now y and z have neighbors y_1 and z_1 respectively in Γ_1. If both y_1 and z_1 were in $\Gamma_1-\Gamma_3$, then the set $K=(I-\{y_1, z_1\}) \cup \{y, z\}$ would be a maximum independent set with $G_3(K)=G_3(I) \cup \{x\}$, which would contradict the maximality of $\alpha_3(I)$. Thus, at least one of y_1 and z_1 is in $\Gamma_1 \cap \Gamma_3$. In this manner we may associate with the vertex v of $\Gamma_1-\Gamma_3$ a vertex of $\Gamma_1 \cap \Gamma_3$. Any such association will be one-to-one as may be verified readily. It follows that $|\Gamma_1-\Gamma_3| \leq |\Gamma_1 \cap \Gamma_3|$. $\blacksquare$

COROLLARY 5. *If G is cubic and contains no 3-cycle, and if I is a maximum independent set with $\alpha_3(I)$ maximum, then*

i) $\alpha_1 \leq 6\alpha_3$; Fajtlowicz [4]
ii) *if G has no 5-cycle,* $\alpha_1 \leq 4\alpha_3$.

Proof. i) Clearly $|\Gamma_1 \cap \Gamma_3| \leq 3\alpha_3$, and $\alpha_1 = |\Gamma_1 - \Gamma_3| + |\Gamma_1 \cap \Gamma_3|$.

ii) By Lemma 2, each vertex in G_3 has at least one neighbor in $I - \Gamma_1$, and thus at most two neighbors in $\Gamma_1 \cap \Gamma_3$. Hence $|\Gamma_1 \cap \Gamma_3| \leq 2\alpha_3$. ■

LEMMA 6. *If G is cubic and contains no 3-cycle and no 5-cycle, then* $\alpha_1 \leq \frac{1}{2}\alpha_2 + \alpha_3$.

Proof. Note that Γ_1 may be split into three disjoint classes: A consisting of vertices with a G_1 neighbor and two G_2 neighbors; B consisting of vertices with a G_1 neighbor, a G_2 neighbor, and a G_3 neighbor; and C consisting of vertices with a G_1 neighbor and two G_3 neighbors. Thus $|A| + |B| + |C| = \alpha_1$. Consider the edge set $[\Gamma_1, G_2]$. The number of edges in $[\Gamma_1, G_2]$ is $2|A| + |B|$, and, by Lemma 2, this number is not bigger than α_2. Thus $2|A| + |B| \leq \alpha_2$. The number of edges in $[\Gamma_1, G_3]$ is $|B| + 2|C|$, and, again by Lemma 2, this number is not bigger than $2\alpha_3$. So $|B| + 2|C| \leq 2\alpha_3$. Adding the two inequalities yields $2|A| + 2|B| + 2|C| \leq \alpha_2 + 2\alpha_3$, or $\alpha_1 \leq \frac{1}{2}\alpha_2 + \alpha_3$. ■

THEOREM 7. *If G is cubic and contains no 3-cycle and no 5-cycle, then* $\alpha_0/n \geq \frac{19}{52}$.

Proof. Recall that $\alpha_0/n = \frac{2}{5} - (\alpha_1 - \alpha_3)/5n$. If I is a maximum independent set with $\alpha_3(I)$ maximum, we have $\alpha_1 \leq 4\alpha_3$, and so $\alpha_0/n \geq \frac{2}{5} - 3\alpha_3/5n$. And, since $\alpha_1 \leq \frac{1}{2}\alpha_2 + \alpha_3 = \frac{1}{2}(\alpha_2 + 2\alpha_3) = \frac{1}{2}(4\alpha_0 - n)$, we have

$$\frac{\alpha_0}{n} \geq \frac{2}{5} - \frac{\frac{1}{2}(4\alpha_0 - n) - \alpha_3}{5n}, \quad \text{or} \quad \frac{\alpha_0}{n} \geq \frac{5}{14} + \frac{\alpha_3}{7n}.$$

Consider two cases:

(i) if $\alpha_3/n \geq \frac{3}{52}$, then $\alpha_0/n \geq \frac{5}{14} + \frac{1}{7}(\frac{3}{52}) = \frac{19}{52}$; and
(ii) if $\alpha_3/n \leq \frac{3}{52}$, then $\alpha_0/n \geq \frac{2}{5} - \frac{3}{5}(\frac{3}{52}) = \frac{19}{52}$.
Thus, in any case, $\alpha_0/n \geq \frac{19}{52}$.

THEOREM 8. *If G is cubic and contains no 3-cycle, no 5-cycle, and no 7-cycle, then* $\alpha_0/n \geq \frac{20}{53}$.

Proof. Let I be a maximum independent set with $\alpha_3(I)$ maximal. We partition I into seven subsets as follows:

A: vertices with two neighbors in G_2, and one in G_1;
B: vertices with one neighbor in G_1, one in G_2, one in G_3;
C: vertices with one neighbor in G_1, two in G_3;
D: vertices with three neighbors in G_2;

E: vertices with two neighbors in G_2, one in G_3;
F: vertices with two neighbors in G_3, one in G_2;
H: vertices with three neighbors in G_3.

We let a, b, c, d, e, f, and h be the respective cardinalities of these sets. Thus,

(1) $a+b+c+d+e+f+h=\alpha_0$
(2) $a+b+c=\alpha_1$
(3) $2a+b+3d+2e+f=2\alpha_2$
(4) $b+2c+e+2f+3h=3\alpha_3$.

Now, let v be a vertex in D. All three of the neighbors of v are in G_2; call them x, y, and z. We claim that not more than one of these may have a neighbor in Γ_1. For, suppose that x has a neighbor x_1 in Γ_1, and that y has a neighbor y_1 in Γ_1. Let x_2 and y_2 be the respective G_1 neighbors of x_1 and y_1. See Fig. 1.

The set $J=(I-\{v, x_1, y_1\})\cup\{x, x_2, y, y_2\}$ is an independent set larger than I, which is a contradiction. Therefore, each vertex in D has at least two G_2 neighbors with no Γ_1 neighbors. A similar argument shows that each vertex in E has at least one G_2 neighbor with no Γ_1 neighbor. Consider the edge set $[\Gamma_1, G_2]$, which has exactly $2a+b$ edges. From Lemma 2, we know that each vertex in G_2 is incident with at most one edge in $[\Gamma_1, G_2]$. But some vertices in G_2 contribute no edges at all to $[\Gamma_1, G_2]$. In particular, there are at least $\frac{1}{2}(2d+e)$ vertices in G_2 which do not contribute to $[\Gamma_1, G_2]$. Thus $2a+b\leq \alpha_2-\frac{1}{2}(2d+e)$, or $4a+2b+2d+e\leq 2\alpha_2$, and, invoking equation (3),

$$4a+2b+2d+e\leq 2a+b+3d+2e+f$$

or $2a+b\leq d+e+f$. But note that $\alpha_0-\alpha_1=d+e+f+h$. Hence $2a+b\leq \alpha_0-\alpha_1$. From the proof of Lemma 6, we have $b+2c\leq 2\alpha_3$. Hence, adding, we obtain

$$2a+2b+2c\leq \alpha_0-\alpha_1+2\alpha_3,$$

or

$$2\alpha_1\leq \alpha_0-\alpha_1+2\alpha_3,$$

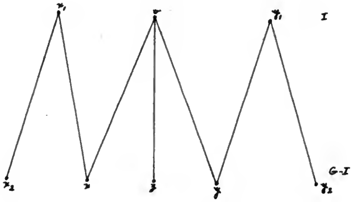

Figure 1

or
$$\alpha_1 \le \tfrac{1}{3}\alpha_0 + \tfrac{2}{3}\alpha_3.$$

Recalling that $\alpha_0/n = \frac{2}{5} - (\alpha_1 - \alpha_3)/5n$, we get

$$\frac{\alpha_0}{n} \ge \frac{2}{5} - \frac{\frac{1}{3}\alpha_0 - \frac{1}{3}\alpha_3}{5n},$$

or

$$\frac{16}{15}\frac{\alpha_0}{n} \ge \frac{2}{5} + \frac{\alpha_3}{15n},$$

or

$$\frac{\alpha_0}{n} \ge \frac{3}{8} + \frac{\alpha_3}{16n}. \tag{5}$$

Recall that in the proof of Theorem 7, we obtained

$$\frac{\alpha_0}{n} \ge \frac{2}{5} - \frac{3\alpha_3}{5n}. \tag{6}$$

Consider two cases:

(i) if $\alpha_3/n \ge \frac{2}{53}$, we use (5) to get $\alpha_0/n \ge \frac{3}{8} + \frac{1}{16}(\frac{2}{53}) = \frac{160}{424} = \frac{20}{53}$;
(ii) if $\alpha_3/n \le \frac{2}{53}$, we use (6) to get $\alpha_0/n \ge \frac{2}{5} - \frac{3}{5}(\frac{2}{53}) = \frac{100}{265} = \frac{20}{53}$.
Thus, in any case, $\alpha_0/n \ge \frac{20}{53}$. ■

We turn our attention at this point to cubic graphs with large girth. If u_j is the greatest lower bound for independence ratios of cubic graphs with girth at least j, then one may ask for $\lim_{j\to\infty} u_j$. We will show this limit is at least $\frac{7}{18}$.

LEMMA 9. *If G is cubic and has girth bigger than $4k+1$, then*

$$\alpha_1 \le \frac{\alpha_2}{k} + f + c - a.$$

Proof. Let l be an edge in $[\Gamma_1, G_2]$. Let v be the end vertex of l in Γ_1, and let w_1 be the end vertex of l in G_2. Let u be the neighbor of v in G_1. Now let $y_1 \ne v$ be the other neighbor of w_1 in I. By Lemma 2, y_1 is not in Γ_1, so y_1 is in $D \cup E \cup F$. So either y_1 has another neighbor $w_2 \ne w_1$ in G_2, or y_1 is in F. If y_1 is in F, we stop. If y_1 is not in F, we consider w_2 and its neighbor $y_2 \ne y_1$ in I. Imitating the proof of Lemma 2, one may show that y_2 is not in Γ_1. So again, either y_2 is in F, or y_2 has a neighbor $w_3 \ne w_2$ in G_2. Continuing in this manner, we obtain, after at most k steps, either a vertex z in F or a sequence w_1, $w_2, \ldots, w_k$ of vertices in G_2. Hence, to each of the $2a+b$ edges in $[\Gamma_1, G_2]$, we may assign either a vertex z in F or a sequence $w_1, \ldots, w_k$ of distinct vertices in G_2.

Suppose that, beginning with l in $[\Gamma_1, G_2]$ we obtain $v, w_1, w_2, \ldots, w_i, z$, with $z \in F$, each w in G_2. And, beginning with another edge l' in $[\Gamma_1, G_2]$

suppose that we obtain v', w_1', $w_2', \ldots, w_j'$, z', with z' in F, and each w' in G_2. If $z=z'$, then $w_i=w_j'$, since z and z' have only one edge into G_2. Hence $y_{i-1}=y_{j-1}'$. Let t be the smallest integer such that $y_t=y_m'$ for some m. The vertices $u, w_1, \ldots, w_t, w_m', w_{m-1}', \ldots, w_1', u'$ are independent, since the girth is at least $4k+2$. Then let $J=(I-\{v, y_1, \ldots, y_t=y_m', y_{m-1}', \ldots, y_1', v'\})\cup\{u, w_1, \ldots, w_t, w_m', w_{m-1}', \ldots, w_1', u'\}$ is an independent set larger than I. This contradiction shows that $z\neq z'$. In a similar manner, one may show that if beginning with l we obtain $w_1, w_2, \ldots, w_k$ in G_2, and if beginning with l' we obtain $w_1', w_2', \ldots, w_k'$ in G_2, then the sets $\{w_1, w_2, \ldots, w_k\}$ and $\{w_1', w_2', \ldots, w_k'\}$ are disjoint. It follows that $[\Gamma_1, G_2]$ has no more than $(1/k)\alpha_2+f$ edges. Hence $2a+b\leq(1/k)\alpha_2+f$. Adding c to both sides of this inequality yields $2a+b+c=\alpha_1+a\leq(1/k)\alpha_2+f+c$. ■

THEOREM 10. *There exists a sequence* $\{\varepsilon_k\}$ *of positive numbers converging to zero such that if* G *is a cubic graph with girth bigger than* $4k+1$, *then the independence ratio* $\alpha_0/n\geq\frac{7}{18}-\varepsilon_k$.

Proof. By Lemma 9, $\alpha_1+a\leq(1/k)\alpha_2+f+c$. Adding $b+c$ to both sides yields

$$2\alpha_1\leq\frac{1}{k}\alpha_2+b+2c+f. \tag{7}$$

Equation (4) from the proof of Theorem 8 says $3\alpha_3=b+2c+e+2f+3h$, so $3\alpha_3-f\geq b+2c+f$. Substituting this inequality into (7) yields

$$2\alpha_1\leq\frac{1}{k}\alpha_2+3\alpha_3-f. \tag{8}$$

Now, from the proof of Lemma 6, we take the inequality $b+2c\leq2\alpha_3$, and substitute it into (7), yielding

$$2\alpha_1\leq\frac{1}{k}\alpha_2+2\alpha_3+f. \tag{9}$$

Adding (8) and (9), and dividing by four gives

$$\alpha_1\leq\frac{1}{2k}\alpha_2+\frac{5}{4}\alpha_3. \tag{10}$$

Recall that $\alpha_0/n=\frac{2}{5}-(\alpha_1-\alpha_3)/5n$. Substituting (10) into this equation we get

$$\frac{\alpha_0}{n}\geq\frac{2}{5}-\frac{1}{10nk}\alpha_2-\frac{1}{20n}\alpha_3.$$

Let $\varepsilon_k=1/18k$. Since $\alpha_2/n\leq1$, we have

$$\frac{\alpha_0}{n}\geq\frac{2}{5}-\frac{1}{10k}-\frac{1}{20}\frac{\alpha_3}{n}. \tag{11}$$

Recall from the proof of Theorem 8 that we have

$$\frac{\alpha_0}{n} \geq \frac{3}{8} + \frac{1}{16}\frac{\alpha_3}{n}.$$

Consider two cases. If $\alpha_3/n \geq \frac{2}{9} - 8/9k$, then (5) yields

$$\frac{\alpha_0}{n} \geq \frac{3}{8} + \frac{1}{16}\left(\frac{2}{9} - \frac{8}{9k}\right) = \frac{3}{8} + \frac{1}{72} - \frac{1}{18k}$$
$$= \frac{7}{18} - \frac{1}{18k} = \frac{7}{18} - \varepsilon_k.$$

If $\alpha_3/n \leq \frac{2}{9} - 8/9k$, then (11) yields

$$\frac{\alpha_0}{n} \geq \frac{2}{5} - \frac{1}{10k} - \frac{1}{20}\left(\frac{2}{9} - \frac{8}{9k}\right)$$
$$= \frac{2}{5} - \frac{1}{90} - \frac{1}{10k} + \frac{8}{180k} = \frac{7}{18} - \varepsilon_k. \quad \blacksquare$$

REFERENCES

1. M. Albertson, B. Bollobas and S. Tucker, *The independence ratio and maximum degree of a graph*, Proc. 7th S-E Conf. Combinatorics, Graph Theory, and Computing, LSU, (1976), 43–50.

2. R. L. Brooks, *On colouring the nodes of a network*, Proc. Cambridge Philos. Soc. **37** (1941), 194–197.

3. P. Erdös, *Graph theory and probability, Canadian J. Math.* **11** (1959), 34–38.

4. S. Fajtlowicz, *The independence ratio for cubic graphs*, Proc. 8th S-E Conf. Combinatorics, Graph Theory, and Computing, LSU, (1977), 273–277.

5. S. Fajtlowicz, *On the size of independent sets in graphs*, Proc. 9th S-E Conf. Combinatorics, Graph Theory, and Computing, Florida Atlantic University, (1978), 269–274.

6. W. Staton, *Independence in graphs with maximum degree three*, Proc. 8th S-E Conf. Combinatorics, Graph Theory, and Computing, LSU, (1977), 615–617.

7. W. Staton, *Some Ramsey-type numbers and the independence ratio*, Trans. Amer. Math. Soc. **256**, (1979), 353–370.

THE UNIVERSITY OF MISSISSIPPI
UNIVERSITY, MISSISSIPPI, 38677

Canad. Math. Bull. Vol. **25** (2), 1982

SOLVABLE GROUPS OF UNIPOTENT ELEMENTS IN A RING

BY
ABRAHAM A. KLEIN

ABSTRACT. Let R be a ring with 1 whose nil subrings are nilpotent modulo the sum of nilpotent ideals. It is proved that if G is a locally solvable group of unipotent elements in R, then the subring generated by $\{g-1 \mid g \in G\}$ is nil. This result implies a result of Sizer showing that a solvable group of unipotent matrices over a skew field can be simultaneously triangularized.

An element x in a ring with 1 is said to be *unipotent* if $x-1$ is nilpotent. A well-known theorem of Kolchin [3] states that a multiplicative semigroup of $n \times n$ unipotent matrices over a field can be put in simultaneous triangular form. The corresponding question over a skew field (infinitely dimensional over its center) is still open. Kaplansky [1] suggests considering a more general question: Let S be a multiplicative semigroup of unipotent elements. Does the set $\{x-1 \mid x \in S\}$ generate a nil subring?

Sizer [6] has considered a group of unipotent matrices over a skew field and has proved that if such a group is solvable then it can be simultaneously triangularized. Following Sizer we consider Kaplansky's question in the case S is a solvable group and we prove that the answer is positive for rings whose nil subrings are nilpotent modulo the nil radical. In particular, the answer is positive for rings whose nil subrings are nilpotent modulo the sum of nilpotent ideals. For a large class of rings in which the last property holds, see Rowen [5, Theorem 6].

A ring R is said to be of *bounded index* n if any nilpotent element $x \in R$ satisfies $x^n = 0$. If the ring of $n \times n$ matrices $M_n(R)$ over a ring R with 1 is of bounded index n, then by [2, Theorem 10] its nil subrings are nilpotent, so our result applies to $M_n(R)$. In particular, it applies to $M_n(D)$, D a skew field, [4], and Sizer's result follows since a nilpotent subring of $M_n(D)$ can be put in simultaneous triangular form.

The following notations and remarks will be used throughout the paper.

Given a multiplicative group, G, of elements of a ring R, we denote the set $\{g-1 \mid g \in G\}$ by $G-1$. If H is a normal subgroup of G, then given $g \in G$,

---

Received by the editors March 24, 1980 and, in revised form, October 10, 1980.
AMS subject classification 1980: Primary 16A22; Secondary 16A42.

$h \in H$ there exists $h' \in H$ such that $hg = gh'$, so $(h-1)g = g(h'-1)$. It follows that if S is the semigroup generated by $H-1$, then $GS = \{gs \mid g \in G,\ s \in S\}$ is the semigroup generated by $G(H-1) = \{g(h-1) \mid g \in G,\ h \in H\}$. Moreover, if U is the subring generated by S, then $GU = \{\sum g_i s_i \mid g_i \in G,\ s_i \in S\}$ is the subring generated by GS. It also follows that if U is nilpotent modulo an ideal of R, then the same is true for GU.

If V, GV denote the subrings generated by $G-1, G(G-1)$ respectively, then the ring GU, which is clearly an ideal in the subring generated by G, is an ideal in the ring GV. Note that GV is in fact the additive subgroup generated by $G(G-1)$.

The nil radical of a ring R will be denoted by $\mathrm{Nil}(R)$. We shall not assume that the group G is solvable, but merely that it is *locally solvable*, which means that every finitely generated subgroup is solvable. Our main result is:

THEOREM. *Let R be a ring whose nil subrings are nilpotent modulo* $\mathrm{Nil}(R)$. *If G is a locally solvable group of unipotent elements of R, then $G(G-1)$ (and in particular $G-1$) generates a nil subring.*

Proof. We have to prove that any given element of the subring generated by $G(G-1)$ is nilpotent. In the expression of such an element, only a finite number of elements of G occur. These elements generate a solvable subgroup since G is locally solvable. This shows that the result of the theorem will follow if it can be proved in the case G is solvable. So let us assume G is solvable and we prove the result by induction on the solvable length l of G.

If $l = 1$, then G is commutative, so the elements of $G(G-1)$ are nilpotent and it follows that $G(G-1)$ generates a nil subring.

Let $G^{(l+1)} = \{1\}$, so by the induction hypothesis applied to $H = G'$ we have that $H-1$ generates a nil subring U, which is nilpotent modulo $\mathrm{Nil}(R)$ by our assumption on R. It follows, by one of the above remarks, that GU is nilpotent modulo $\mathrm{Nil}(R)$ so GU is nil. If as above V denotes the subring generated by $G-1$ then GU is an ideal in GV, so the result of the theorem will follow if we prove that GV/GU is nil.

We first prove that GV/GU is commutative. If $g_1, g_2 \in G$ then $g_1g_2 - g_2g_1 = g_2g_1(g_1^{-1}g_2^{-1}g_1g_2 - 1) \in GU$. This implies that the subring generated by G is commutative $\bmod GU$, so also GV is commutative $\bmod GU$. To prove that GV/GU is nil, we recall that GV is the additive subgroup generated by $G(G-1)$, so it suffices to show that if $g \in G$ and $t \in G-1$ then gt is nilpotent $\bmod GU$. But this follows by commutativity $\bmod GU$ and by the assumption that the elements of G are unipotent. With this the theorem has been proved.

The result of the theorem together with [2, Theorem 10] yield:

COROLLARY 1. *Let R be a ring with 1 and let $M_n(R)$ be of bounded index n. If G is a solvable group of unipotent matrices of $M_n(R)$, then the subring generated by $G(G-1)$ (so also that generated by $G-1$) is nilpotent.*

Note that the above corollary can be obtained from a more general one, which states that if R is a ring of bounded index and G is a locally solvable group of unipotent elements of R, then $G(G-1)$ generates a ring which is nilpotent modulo the sum of nilpotent ideals. This more general result holds since the class of rings which is given in [5, Theorem 6] includes the rings of bounded index. It has been pointed out in [5] that if R is a ring with 1 and $M_n(R)$ is of bounded index n then $M_n(R)$ has no nonzero nilpotent ideals, so it follows that $G(G-1)$ generates a nilpotent subring.

The following corollary follows from the previous one and the fact that a nilpotent ring of matrices over a skew field can be simultaneously triangularized.

COROLLARY 2 (Sizer [6]). *Let D be a skew field and G a solvable group of unipotent matrices of $M_n(D)$. Then there exists an invertible matrix $P \in M_n(D)$ such that $P^{-1}AP$ is triangular for all $A \in G$.*

The above corollaries have been stated for G solvable rather than for G locally solvable since in the rings considered, if G is locally solvable then it is necessarily solvable. This result follows from the following:

PROPOSITION 1. *Let G be a group of unipotent elements of a ring R. If $G-1$ generates a nilpotent semigroup (subring) then G is solvable.*

Proof. Assume that for some $r \geq 1$ we have $t_1 \cdots t_r = 0$ for all $t_1, \ldots, t_r \in G-1$. We prove by induction on r that $G^{(r-1)} = \{1\}$.

If $r=1$ the result is trivial. If $r \geq 2$, let $H = G'$ and we prove that $H-1$ generates a nilpotent semigroup of index $\leq r-1$ and this will imply $H^{(r-2)} = \{1\}$ by the induction hypothesis, so $G^{(r-1)} = \{1\}$.

In fact we show that $t_1 \cdots t_{r-2}s = 0$ for all $t_1, \ldots, t_{r-2} \in G-1$ and $s \in H-1$. First we show that this holds for $s = h-1$, $h = g_1g_2g_1^{-1}g_2^{-1}$ a commutator of any two elements $g_1, g_2 \in G$. Let $t_{r-1} = g_1^{-1}$, $t_r = g_2 - 1$ then by our assumption on G we have

$$t_1 \cdots t_{r-2}t_{r-1}t_r = 0 = t_1 \cdots t_{r-2}t_rt_{r-1}.$$

So

$$t_1 \cdots t_{r-2}(g_1g_2 - g_2g_1) = t_1 \cdots t_{r-2}(t_{r-1}t_r - t_rt_{r-1}) = 0$$

and this implies that $t_1 \cdots t_{r-2}(g_1g_2g_1^{-1}g_2^{-1} - 1) = 0$. Now if $s \in H-1$ then $s = h_1 \cdots h_q - 1$ where $h_1, \ldots, h_q$ are commutators of elements of G. We may assume by induction that $t_1 \cdots t_{r-2}(h_1 \cdots h_{q-1} - 1) = 0$, so we obtain

$$t_1 \cdots t_{r-2}s = t_1 \cdots t_{r-2}((h_1 \cdots h_{q-1} - 1)h_q + h_q - 1) = 0.$$

The above proposition together with corollary 1 yield:

PROPOSITION 2. *Let R be a ring such that $M_n(R)$ has bounded index n. Let G*

be a group of unipotent matrices of $M_n(R)$. *Then the semigroup generated by* $G-1$ *is nilpotent if and only if* G *is solvable.*

We now go back to the theorem and make a weaker assumption on the subrings of the ring R. We assume that they are locally nilpotent modulo Nil(R). A positive result can still be obtained if we make a stronger assumption on G. We only state the result. Its proof is clear by our previous arguments.

PROPOSITION 3. *Let* R *be a ring whose nil subrings are locally nil-potent modulo* Nil(R) *and let* G *be a group of unipotent elements of* R. *If* G *is locally finite and locally solvable then* $G(G-1)$ *generates a nil subring.*

We conclude with a result which follows from Proposition 1.

PROPOSITION 4. *Let* G *be a locally finite group of unipotent elements of a ring* R. *If* $G-1$ *generates a locally nilpotent semigroup* (*subring*), *then* G *is locally solvable.*

REFERENCES

1. I. Kaplansky, *The Engel-Kolchin theorem revisited, Contributions to Algebra* (*collection of papers dedicated to Ellis Kolchin*) 233–237. Academic Press, 1977.

2. A. A. Klein, *Some ring-theoretic properties implied by embeddability in fields,* Journal of Algebra (to appear).

3. E. Kolchin, *On certain concepts in the theory of algebraic matric groups,* Ann. of Math. **49** (1948), 774–789.

4. J. Levitzki, *Über nilpotente Unterringe,* Math. Ann. **105** (1931), 620–627.

5. L. H. Rowen, *Rings and semigroups satisfying quasi-identities,* preprint.

6. W. S. Sizer, *Triangularizing solvable groups of unipotent matrices over a skew field,* Canad. Math. Bull. **20** (1977), 121–123.

DEPARTMENT OF MATHEMATICAL SCIENCES
TEL-AVIV UNIVERSITY
TEL-AVIV, ISRAEL

Canad. Math. Bull. Vol. **25** (2), 1982

NOTES ON THE BIRKHOFF ALGORITHM FOR DOUBLY STOCHASTIC MATRICES

BY
RICHARD A. BRUALDI

ABSTRACT. The purpose of this note is to tie together some results concerning doubly stochastic matrices and their representations as convex combinations of permutation matrices.

Let Ω_n denote the set of $n\times n$ *doubly stochastic* matrices. Thus an $n\times n$ matrix $S=[s_{ij}]$ belongs to Ω_n provided $s_{ij}\geq 0$ $(i,j=1,\ldots,n)$, and the sum of the entries in each row and in each column of S equals one. It is readily verified that Ω_n is a convex polytope in Euclidean n^2-space, and it was shown in [8] that Ω_n has dimension n^2-2n+1. Birkhoff [1; 10, p. 58] proved that the extreme points of Ω_n are the $n\times n$ permutation matrices, and in doing so gave an algorithm for writing a doubly stochastic matrix as a convex combination of permutation matrices.

Let $A=[a_{ij}]$ be an $n\times n$ matrix of 0's and 1's such that for each (e,f) with $a_{ef}=1$, there is a permutation matrix $P=[p_{ij}]$ where $p_{ef}=1$ and $P\leq A$ (the order relation is entrywise). Then A is said to have *total support.* It was proved in [3] that the faces of Ω_n are in one-to-one correspondence with the $n\times n$ matrices A of 0's and 1's with total support; the face of Ω_n associated with A is

$$\mathcal{F}(A)=\{S\in\Omega_n : S\leq A\}.$$

If A has total support, then there is a largest integer $k\geq 1$ such that after reordering rows and columns, A is a direct sum $A_1\oplus\cdots\oplus A_k$ where $A_1,\ldots,A_k$ have total support. Since k is maximum $A_1,\ldots,A_k$ cannot be decomposed further into direct sums and are called the *fully indecomposable components* of A. It was proved in [3] by an elementary combinatorial argument that

$$\dim\mathcal{F}(A)=\sigma(A)-2n+k$$

where $\sigma(A)$ is the number of positive entries of A. It now follows from the

---

Received by the editors April 29, 1980.

AMS Subject Classification Numbers: 15A51, 05C50, 52A25.

Research partially supported by National Science Foundation Grant No. MCS 76-06374 A01.

classical theorem of Caratheodory [5] that every matrix in $\mathcal{F}(A)$ can be written as a convex combination of no more than $\sigma(A)-2n+k+1$ permutation matrices.

Let $S=[s_{ij}]\in\Omega_n$ and let $S_{(0,1)}$ be the matrix obtained from S by replacing each positive entry with a one. Thus $\mathcal{F}(S_{(0,1)})$ is the smallest face of Ω_n containing S. The *Birkhoff algorithm* for expressing S as a convex combination of permutation matrices proceeds as follows. Choose a smallest positive entry $c_1=s_{kl}$ of S. Then there exists a permutation matrix $P_1=[p_{ij}]$ such that $p_{kl}=1$ and $P_1\le S_{(0,1)}$. The matrix $S_1=S-c_1P_1$ is a nonnegative matrix having at least one more zero entry than S. We repeat with S_1 replacing S (the matrix S_1, if not the zero matrix, can be normalized to a doubly stochastic matrix if desired by multiplying by $1/(1-c_1)$), and continue until the zero matrix is obtained. We shall relax this algorithm somewhat by reversing the order of its two major components. Namely, we first choose a permutation matrix $P_1\le S_{(0,1)}$ and then choose the smallest (positive) entry in those positions of S corresponding to the 1's of P. The matrix $S_1=S-c_1P_1$ still has at least one more zero entry than S. We call any representation of S as a convex combination $c_1P_1+\cdots+c_tP_t$ of permutation matrices which is obtained in this way a *Birkhoff representation* of S.

How good is the Birkhoff algorithm? Good enough so that it expresses S as a convex combination of permutation matrices whose number does not exceed that which is guaranteed by Caratheodory's theorem:

(1) Birkhoff's algorithm expresses the doubly stochastic matrix S as a convex combination of no more than $\sigma(S_{(0,1)})-2n+k+1$ permutation matrices (where k is the number of fully indecomposable components of $S_{(0,1)}$).

The property (1) above was first established by Johnson, Dulmage, and Mendelsohn [7]. Another proof has been given by Nishi [9]. An alternative, and perhaps more revealing, proof can be given using the property:

(2) Each step of Birkhoff's algorithm lowers by at least one the dimension of the smallest face of Ω_n containing the doubly stochastic matrix.

To verify (2), let S be an $n\times n$ doubly stochastic matrix. It suffices to assume that $S_{(0,1)}$ is fully indecomposable. Let $T=(1-c)^{-1}\,(S-cP)$ where $P\le S_{(0,1)}$ is a permutation matrix and c is the smallest positive entry of S at positions corresponding to the 1's of P. Then T is doubly stochastic and has at least one more zero entry than S. The matrix $T_{(0,1)}$ then has total support and at least one more zero entry than $S_{(0,1)}$. Suppose that $T_{(0,1)}$ has $p\ge1$ fully indecomposable components. Then after reordering of rows and columns, $T_{(0,1)}=B_1\oplus\cdots\oplus B_p$ where $B_1,\ldots,B_p$ are fully indecomposable. Since $T_{(0,1)}\le S_{(0,1)}$

and $S_{(0,1)}$ is fully indecomposable, it follows that $\sigma(S_{(0,1)}) \geq \sigma(T_{(0,1)}) + p$. Hence

$$\begin{aligned} \dim \mathcal{F}(T_{(0,1)}) &= \sigma(T_{(0,1)}) - 2n + p \\ &\leq \sigma(S_{(0,1)}) - p - 2n + p \\ &\leq \dim \mathcal{F}(S_{(0,1)}) - 1. \end{aligned}$$

Property (1) is now a consequence of (2). For, if $S \in \Omega_n$ and $S_{(0,1)}$ has k fully indecomposable components, then $\dim \mathcal{F}(S_{(0,1)}) = \sigma(S_{(0,1)}) - 2n + k$. Hence after at most $\sigma(S_{(0,1)}) - 2n + k$ steps of the Birkhoff algorithm, the resulting doubly stochastic matrix lies in a face of Ω_n of dimension 0, that is, is a permutation matrix. It follows that S is then expressed as a convex combination of at most $\sigma(S_{(0,1)}) - 2n + k + 1$ permutation matrices.

A lower bound for the minimum number of permutation matrices in a representation of a doubly stochastic matrix as a convex combination of permutation matrices can be obtained as follows. Let $A = [a_{ij}]$ be an $n \times n$ matrix of 0's and 1's with total support. Let U be a set of positions of A such that $(k, l) \in U$ implies $a_{kl} = 1$. Then U is said to be *strongly stable* [2] provided each permutation matrix $P = [p_{ij}]$ with $P \leq A$ satisfies $p_{kl} = 1$ for at most one $(k, l) \in U$. Denote by $\alpha(A)$ the maximum cardinality of a strongly stable set of positions of A. In [2] it is shown that $\alpha(A) = \max \sigma(X)$ where the maximum is taken over all submatrices X of A such that the submatrix of A complementary to X (the one formed by the rows complementary to the rows in which X lies and the columns complementary to the columns in which X lies) is a $p \times (n-1-p)$ zero submatrix for some $p = 0, 1, \ldots, n-1$. In particular, $\alpha(A)$ is at least equal to the maximum number of 1's in a row (take $p = n-1$) and in a column (take $p = 0$) of A. The following property is immediate.

(3) The $n \times n$ doubly stochastic matrix S cannot be expressed as a convex combination of fewer than $\alpha(S_{(0,1)})$ permutation matrices. In particular, Birkhoff's algorithm expresses S as a convex combination of at least $\alpha(S_{(0,1)})$ permutation matrices.

In (1) and (3) are given upper and lower bounds for the number of steps required in any application of Birkhoff's algorithm to a doubly stochastic matrix. We shall show that in each case equality can always be obtained and in the following strong sense: no matter how Birkhoff's algorithm is carried out the number of permutation matrices achieves equality. Before doing so, we first illustrate the following property:

(4) The number of permutation matrices in a Birkhoff representation of a doubly stochastic matrix S need not be unique.

Let S be the doubly stochastic matrix

$$\frac{1}{6}\begin{bmatrix}1&4&0&1\\2&1&3&0\\2&1&1&2\\1&0&2&3\end{bmatrix}.$$

Then $S=\frac{1}{6}P_1+\frac{1}{6}P_2+\frac{1}{3}P_3+\frac{1}{3}P_4$ is a Birkhoff representation of S, where

$$P_1=\begin{bmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix},\quad P_2=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},$$

$$P_3=\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix},\quad P_4=\begin{bmatrix}0&1&0&0\\0&0&1&0\\1&0&0&0\\0&0&0&1\end{bmatrix}.$$

Also $S=\frac{1}{6}Q_1+\frac{1}{6}Q_2+\frac{1}{6}Q_3+\frac{1}{6}Q_4+\frac{1}{3}Q_5$ is a Birkhoff representation of S where

$$Q_1=\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix},\quad Q_2=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix},\quad Q_3=\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},$$

$$Q_4=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix},\quad Q_5=\begin{bmatrix}0&1&0&0\\0&0&1&0\\1&0&0&0\\0&0&0&1\end{bmatrix}.$$

Let A be an $n\times n$ matrix of 0's and 1's with total support. Then from the formula for $\alpha(A)$ it follows that the strong stability number $\alpha(A)$ of A satisfies $1\leq\alpha(A)\leq\lceil n/2\rceil\lceil(n+1)/2\rceil$. If $n\geq 2$ and A is fully indecomposable, then each row and column of A contains at least two 1's, so that $2\leq\alpha(A)$.

(5) Let t be an integer with $1\leq t\leq\lceil n/2\rceil\lceil(n+1)/2\rceil$. Then there exists an $n\times n$ doubly stochastic matrix S such that $\alpha(S_{(0,1)})=t$ and each Birkhoff representation of S uses exactly t distinct permutation matrices.

If $t=1$, then in (5) we may take S to be an $n\times n$ permutation matrix. To complete the verification we let $t\geq 2$ and construct an $n\times n$ fully indecomposable matrix A of 0's and 1's with $\alpha(A)=t$ such that $\mathscr{F}(A)$ is a simplex of dimension $t-1$. It then follows that a doubly stochastic matrix S in the interior of $\mathscr{F}(A)$ (that is, $S_{(0,1)}=A$) requires exactly t distinct permutation matrices in each of its representations as a convex combination of permutation matrices.

Let p and q be positive integers with $p+q=n-1$. Let E be the $(q+1)\times q$ vertex-edge incidence matrix of a tree and F the $(p+1)\times p$ vertex-edge incidence matrix of a tree. Let X be any $(q+1)\times(p+1)$ matrix of 0's and 1's such that the $n\times n$ matrix

$$A=\begin{bmatrix} X & E \\ F^t & 0 \end{bmatrix}$$

has at least two 1's in each row and column. Then it follows from results in [3] that A is fully indecomposable and $\mathscr{F}(A)$ is a simplex of dimension $\sigma(A)-2n+1=\sigma(X)-1$. Since $\alpha(A)\geq\sigma(X)$, it follows that $\alpha(A)=\sigma(X)$. Now choose $p=\lfloor(n-1)/2\rfloor$ and $q=\lfloor n/2\rfloor$, and choose E and F to be vertex-edge incidence matrices of paths whose initial and terminal vertices correspond, respectively, to the first and last row. Then all row and column sums of A will be at least 2 if the entries of X in positions $(1, 1)$ and $(q+1, p+1)$ equal one. Since $(q+1)\times(p+1)=\lceil n/2\rceil\lceil(n+1)/2\rceil$, the conclusion now follows.

To construct matrices for which equality occurs in (1), we first need to review some ideas. A fully indecomposable matrix A is called *nearly decomposable* provided each matrix obtained from A by replacing a 1 with 0 is not fully indecomposable (it could not even be of total support). Let A be an $n\times n$ nearly decomposable matrix of total support with $n\geq 2$. Then [6, see also 4] there exists an integer n_1 with $1\leq n_1\leq n-1$ and an $n_1\times n_1$ nearly decomposable matrix A_1 such that after reordering rows and columns, A has the form

$$\left[\begin{array}{c|ccccc} A_1 & & & E & & \\ \hline & 1 & 0 & \cdots & 0 & 0 \\ & 1 & 1 & \cdots & 0 & 0 \\ F & \vdots & \vdots & & \vdots & \vdots \\ & 0 & 0 & \cdots & 1 & 0 \\ & 0 & 0 & \cdots & 1 & 1 \end{array}\right] \tag{6}$$

where all entries of E and F are zero except for the entry of each in the upper right corner. If $n_1>1$, a similar decomposition holds for A_1, and so on.

Associated with an $n\times n$ matrix $A=[a_{ij}]$ of 0's and 1's is a bipartite graph $BG(A)$ of $2n$ vertices $x_1,\ldots,x_n, y_1,\ldots,y_n$; there is an edge joining x_i and y_j if and only if $a_{ij}=1$ $(i, j=1,\ldots,n)$, and these are the only edges. Let A be fully indecomposable. By rearranging rows and columns, we may assume that $I_n\leq A$ where I_n is the $n\times n$ identity matrix. Since A is fully indecomposable, $BG(A)$ is connected. Hence there is a spanning tree of $BG(A)$ which has among its edges the n edges $[x_1, y_1],\ldots,[x_n, y_n]$ of $BG(A)$. This spanning tree is associated with an $n\times n$ matrix T of 0's and 1's where $I_n\leq T\leq A$. We call T a *tree matrix*; note that $\sigma(T)=2n-1$.

For $1 \le p, q \le n$, the $n \times n$ matrix $E = [e_{ij}]$ where $e_{pq} = 1$ and $e_{ij} = 0$ otherwise is denoted by E_{ij}^n.

(7) Let A be an $n \times n$ fully indecomposable matrix of 0's and 1's. Then there is a tree matrix T with $I_n \le T \le A$ such that the $p = \sigma(A) - 2n + 1$ positions of $A - T$ whose entries equal one can be ordered as $(i_1, j_1), \ldots, (i_p, j_p)$ where for $l = 1, \ldots, p$ there is a permutation matrix P_l whose (i_l, j_l) entry equals one and

$$P_l \le T + E_{i_1j_1}^n + \cdots + E_{i_lj_l}^n.$$

We first verify (7) for nearly decomposable matrices by induction on n. The case $n = 1$ being trivial, we let $n \ge 2$. We may assume A has the form (6) where A_1 is an $n_1 \times n_1$ nearly decomposable matrix and $1 \le n_1 \le n - 1$. By induction there is a tree matrix T_1 with $I_{n_1} \le T_1 \le A_1$ such that the s positions of $A_1 - T_1$ with a nonzero entry can be ordered as $(i_1, j_1), \ldots, (i_s, j_s)$ where for $l = 1, \ldots, s$ there is a permutation matrix Q_l whose (i_l, j_l) entry is one and

$$Q_l \le T_1 + E_{i_1j_1} + \cdots + E_{i_lj_l}.$$

Let

$$T = \left[\begin{array}{c|ccccc} T_1 & & & 0 & & \\ \hline & 1 & 0 & \cdots & 0 & 0 \\ & 1 & 1 & \cdots & 0 & 0 \\ F & \vdots & \vdots & & \vdots & \vdots \\ & 0 & 0 & \cdots & 1 & 0 \\ & 0 & 0 & \cdots & 1 & 1 \end{array}\right]$$

Then $I_n \le T \le A$ and T is a tree matrix. Since A is nearly decomposable (and hence fully indecomposable) there is a permutation matrix $P \le A$ whose $(1, n)$-entry equals one. It now follows easily that the ordering $(i_1, j_1), \ldots, (i_s, j_s), (1, n)$ of the positions of $A - T$ with a nonzero entry has the required properties; the required permutation matrices are $P_1 = Q_1 \oplus I_{n-n_1}, \ldots, P_s = Q_s \oplus I_{n-n_1}$, $P_{s+1} = P$. Thus (7) holds for nearly decomposable matrices. Now let B be an $n \times n$ fully indecomposable matrix which is not nearly decomposable. Then by replacing certain 1's with 0's, we arrive at a nearly decomposable matrix A with $A \le B$. Let T be a tree matrix with $I_n \le T \le A$ satisfying (7). Ordering the positions of $B - T$ with nonzero entries as $(i_1, j_1), \ldots, (i_p, j_p), (i_{p+1}, j_{p+1}), \ldots, (i_q, j_q)$ where $(i_{p+1}, j_{p+1}), \ldots, (i_q, j_q)$ are the positions of $B - A$ with a nonzero entry in any order, we complete the verification.

We now construct doubly stochastic matrices for which equality holds in (1).

(8) Let A be an $n\times n$ fully indecomposable matrix of 0's and 1's. Then in the notation of (7) the matrix

$$S=\frac{1}{2^{p+1}-1}(2^pI_n+2^{p-1}P_1+\cdots+2P_{p-1}+P_p)$$

is in the face $\mathcal{F}(A)$ of dimension $p=\sigma(A)-2n+1$. Every representation of S as a convex combination of permutation matrices requires at least $p+1$ distinct permutation matrices. Every Birkhoff representation of S requires exactly $p+1$ permutation matrices.

To verify (8) we argue with $S'=(2^{p+1}-1)S$ instead of S. Suppose S' is written as a nonnegative linear combination of t permutation matrices $Q_1,\ldots,Q_t$. At least one of the matrices $Q_1,\ldots,Q_t$ has a one in the (i_p,j_p)-position, and the sum of the coefficients of all such matrices is $b_p=1$. Hence at least one of the permutation matrices $Q_1,\ldots,Q_t$ has a zero in the (i_p,j_p)-position and a one in the (i_{p-1},j_{p-1})-position; the sum of the coefficients of all such matrices is b_{p-1} where $1\le b_{p-1}\le 2$. Since $b_p+b_{p-1}\le 3$, at least one of the permutation matrices $Q_1,\ldots,Q_t$ has a zero in positions (i_p,j_p), (i_{p-1},j_{p-1}), and a one in position (i_{p-2},j_{p-2}); the sum of the coefficients of all such matrices is b_{p-2} where $1\le b_{p-2}\le 2^2$. Continuing like this we see that $t\ge p+1$. Since $\dim\mathcal{F}(A)=p=\sigma(A)-2n+1$, it follows from (1) that each Birkhoff representation of S requires at most $p+1$ distinct permutation matrices and hence exactly $p+1$ distinct permutation matrices.

Let A be an $n\times n$ matrix of 0's and 1's of total support with k fully indecomposable components. Using (8) one can construct doubly stochastic matrices $S\in\mathcal{F}(A)$ whose Birkhoff representations require exactly $\dim\mathcal{F}(A)+1=\sigma(A)-2n+k+1$ distinct permutation matrices.

We conclude these notes with some questions. Let $m=n!$ Since the vertices of Ω_n are the $n\times n$ permutation matrices $P_1,\ldots,P_m$, the map

$$f:(c_1,\ldots,c_m)\to c_1P_1+\cdots+c_mP_m$$

from the $(m-1)$-dimensional simplex $\{(c_1,\ldots,c_m): c_i\ge 0, \sum_i c_i=1\}$ to Ω_n is a continuous surjection, and for each $S\in\Omega_n$, $f^{-1}(S)$ is a compact convex set. Each m-tuple in $f^{-1}(S)$ furnishes a representation of the doubly stochastic matrix S as a convex combination of permutation matrices. It is a fundamental and very difficult question to ask for the m-tuple in $f^{-1}(S)$ with the fewest number of positive coordinates, that is, for a representation of S as a convex combination of the smallest possible number $c(S)$ of distinct permutation matrices. The properties (1) and (3) furnish upper and lower bounds on $c(S)$. It seems unlikely that a Birkhoff representation always exists with $c(S)$ distinct permutation matrices, although we have not constructed an example to substantiate this. It is easy to see that a Birkhoff representation corresponds to an extreme point of the polytope $f^{-1}(S)$. Are there other extreme points? The answer most likely is in the affirmative; otherwise, there would have to be a Birkhoff representation of S with $c(S)$ permutation matrices.

Doubly stochastic matrices have been used as a basic data structure for storing and processing uncertain information in multisensor data correlation [12, 13]. In these papers, the *entropy* of an $n \times n$ doubly stochastic matrix S has been defined by

$$H(S) = \sup\{h(c_1, \dots, c_m) : (c_1, \dots, c_m) \in f^{-1}(S)\},$$

where $H(c_1, \dots, c_m) = \sum_{i=1}^{m} -c_i \log c_i$, the entropy of the probability distribution $(c_1, \dots, c_m)$. Thus the entropy of a permutation matrix is 0 and that of the $n \times n$ matrix J_n with each entry equal to $1/n$ is $\log n!$ The entropy of every other $n \times n$ doubly stochastic matrix S satisfies $0 \le H(S) \le \log n!$. An $n \times n$ doubly stochastic matrix S can be regarded as a probabilistic model of a bijection $h: X \to Y$ between two sets of n elements, a *fuzzy bijection*. The row sums of S being one means every element of X for sure goes someplace; the column sums of S being one means for sure every element of Y comes from someplace. Birkhoff's theorem then becomes: every fuzzy bijection is a convex combination of bijections. Since one interpretation of entropy of a probability distribution is that of a measure of uncertainty, the entropy of a fuzzy bijection (doubly stochastic matrix) can be interpreted as a measurement of the uncertainty concerning which bijection is involved. For a doubly stochastic matrix S, which representation as a convex combination of permutation matrices has entropy equal to $H(S)$? What is the largest entropy of a Birkhoff representation?

Finally we note that an extension of Birkhoff's algorithm to doubly stochastic matrices over more general number systems has been considered in [11].

REFERENCES

1. G. Birkhoff, *Tres observaciones sobre el álgebra lineal*, Univ. Nac. Tucumań Rev. Ser. A, **5** (1946), 147–150.

2. R. A. Brualdi, *The diagonal hypergraph of a matrix (bipartite graph)*, Discrete Math. **27** (1979), 127–147.

3. R. A. Brualdi and P. M. Gibson, *Convex polyhedra of doubly stochastic matrices. I. Applications of the permanent function*, J. Comb. Theory Ser. A. **22** (1977), 194–230.

4. R. A. Brualdi and M. B. Hedrick, *A unified treatment of nearly reducible and nearly decomposable matrices*, Lin. Alg. and its Applics. **24** (1979), 51–73.

5. B. Grünbaum, *Convex Polytopes*, Interscience, New York, 1967.

6. D. J. Hartfiel, *A simplified form for nearly reducible and nearly decomposable matrices*, Proc. Amer. Math. Soc. **24** (1970), 388–393.

7. D. N. Johnson, A. L. Dulmage, and N. S. Mendelsohn, *On an algorithm of G. Birkhoff concerning doubly stochastic matrices*, Canad. Math. Bull. **3** (1960), 237–242.

8. M. Marcus and R. Ree, *Diagonals of doubly stochastic matrices*, Quart. J. Math. Oxford, Ser. (2), **10** (1959), 296–302.

9. A. Nishi, *An elementary proof of Johnson–Dulmage–Mendelsohn's refinement of Birkhoff's theorem on doubly stochastic matrices*, Canad. Math. Bull. **22** (1979), 81–86.

10. H. J. Ryser, *Combinatorial Mathematics*, Carus Math. Monograph No. 14, Math. Assoc. of America, Washington, D.C., 1963.

11. H. Schneider, *The Birkhoff–Egervary–König theorem for matrices over lattice-ordered groups*, Acta Math. Acad. Sci. Hungar. **30** (1977), 91–94.

12. F. Witte, *Doubly stochastic matrices and sequential data association, Part I, Information linkage between applied mathematics and industry*, 641–646, Academic Press, New York, 1979.

13. F. Witte and D. Lucas, *Probabilistic tracking in a multitarget environment*, I.E.E.E., 1978, 1212–1219.

Department of Mathematics
University of Wisconsin
Madison, WI 53706, U.S.A.

Canad. Math. Bull. Vol. **25** (2), 1982

ON THE DIVISIBILITY OF THE CLASS NUMBERS OF $Q(\sqrt{-p})$ AND $Q(\sqrt{-2p})$ BY 16.

BY

PHILIP A. LEONARD AND KENNETH S. WILLIAMS*

ABSTRACT. Let $h(m)$ denote the class number of the quadratic field $Q(\sqrt{m})$. In this paper necessary and sufficient conditions for $h(m)$ to be divisible by 16 are determined when $m=-p$, where p is a prime congruent to 1 modulo 8, and when $m=-2p$, where p is a prime congruent to ±1 modulo 8.

0. **Introduction.** Let $D=-p$, where p is a prime congruent to 1 modulo 8, or $D=-2p$, where p is a prime congruent to ± 1 modulo 8. Let $h(D)$ denote the class number of the imaginary quadratic field $Q(\sqrt{D})$. For these values of D, the 2-Sylow subgroup $H_2(D)$ of the class group $H(D)$ of $Q(\sqrt{D})$ is cyclic of order ≥ 4, so that $h(D)\equiv 0 \pmod 4$. Moreover, in each of these cases, necessary and sufficient conditions for $h(D)$ to be divisible by 8 are known in terms of congruences involving the positive integers u and v in the representation

$$p=u^2-2v^2. \tag{0.1}$$

In this paper, using the fact that $H_2(D)$ is cyclic, we determine the corresponding criteria for $h(D)$ to be divisible by 16.

1. $D=-p$, $p\equiv 1 \pmod 8$. We set $g=u+v$, $h=u+2v$ so that g and h are odd positive integers satisfying

$$p=2g^2-h^2. \tag{1.1}$$

Clearly we have

$$\text{G.C.D.}\,(g,p)=\text{G.C.D.}\,(h,p)=\text{G.C.D.}\,(g,h)=1. \tag{1.2}$$

Brown [3: Theorem 2] has shown that

$$h(-p)\equiv 0 \pmod 8 \Leftrightarrow \left(\frac{g}{p}\right)=+1, \tag{1.3}$$

Received by the editors May 2, 1980.

AMS (1980) classification numbers: 12A25, 12A50.

Key words and phrases: class number, imaginary quadratic field, binary quadratic forms.

* Research supported by the Natural Sciences and Engineering Research Council of Canada under grant A-7233.

and Hasse [6: p. 168] has shown that

$$h(-p)\equiv 0 \pmod 8 \Leftrightarrow g\equiv 1 \pmod 4. \tag{1.4}$$

It is easy to see that (1.3) and (1.4) are equivalent since, by appealing to (1.1) and the law of quadratic reciprocity, we have

$$\left(\frac{g}{p}\right)=\left(\frac{p}{g}\right)=\left(\frac{2g^2-h^2}{g}\right)=\left(\frac{-h^2}{g}\right)=\left(\frac{-1}{g}\right). \tag{1.5}$$

We prove the following theorem.

THEOREM 1. *Let* $p\equiv 1 \pmod 8$ *be a prime such that* $h(-p)\equiv 0 \pmod 8$. *Set* $p=2g^2-h^2$, *where* g *and* h *are odd positive integers. As* $h(-p)\equiv 0 \pmod 8$ *we have* $(-1/g)=(g/p)=+1$. *Then*

$$h(-p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{g}{p}\right)_4=\left(\frac{2h}{g}\right).$$

Proof. We consider integral positive-definite binary quadratic forms $ax^2+bxy+cy^2$ (written (a, b, c)) of discriminant $b^2-4ac=-4p$. Clearly b must be even. Moreover all such forms are primitive, that is, G.C.D $(a, b, c)=1$. The class A of forms equivalent to the form (a, b, c) under an integral unimodular transformation of determinant $+1$ is written $A=[a, b, c]$. The product A_1A_2 of two such classes A_1 and A_2 is defined as follows: choose forms $(a_1, b, a_2c)\in A_1$ and $(a_2, b, a_1c)\in A_2$ and define A_1A_2 to be $[a_1a_2, b, c]$. These classes, with the multiplication specified above, form a finite abelian group $\mathcal{H}$, which is isomorphic to the class group $H(-p)$ of the imaginary quadratic field $Q(\sqrt{-p})$. Its order is the class number $h(-p)$.

The identity of $\mathcal{H}$ is the class $I=[1, 0, p]$ and the inverse class of $[a, b, c]\in\mathcal{H}$ is $[a, -b, c]$.

Setting $A=[2, 2, \frac{1}{2}(p+1)]\in\mathcal{H}$, $B=[g, 2h, 2g]\in\mathcal{H}$, it is easy to check that

$$B^2=A\neq I, \qquad A^2=I, \tag{1.6}$$

so that

$$\operatorname{ord}(A)=2, \qquad \operatorname{ord}(B)=4. \tag{1.7}$$

As $(g/p)=+1$, the form $(g, 2h, 2g)$ represents an integer s, namely $s=g$, satisfying

$$\left(\frac{-1}{s}\right)=\left(\frac{s}{p}\right)=+1, \qquad (s, 2p)=1.$$

Thus B belongs to the principal genus of $\mathcal{H}$, and so, by Gauss' duplication theorem, is the square of some class $C=[l, m, n]$, that is,

$$C^2=B. \tag{1.8}$$

Clearly we have

(1.9) $$\operatorname{ord}(C)=8.$$

Replacing (l, m, n) by an equivalent form, we can suppose that

(1.10) $$\text{G.C.D.}(l, 2gp)=1.$$

We will now show that

(1.11) $$h(-p)\equiv 0 (\bmod 16) \Leftrightarrow \left(\frac{l}{p}\right)=+1.$$

If $(l/p)=+1$ then, as C represents l, C must belong to the principal genus, and so is the square of some class D. From (1.9) we have $\operatorname{ord}(D)=16$, and so $h(-p)\equiv 0 \pmod{16}$.

Conversely if $h(-p)\equiv 0 \pmod{16}$, since the 2-Sylow subgroup of $\mathscr{H}$ is cyclic by a theorem of Gauss, $\mathscr{H}$ contains an element D of order 16. Thus D^2 is of order 8. But there are exactly four elements of order 8 in $\mathscr{H}$, namely C, C^3, C^5, C^7, thus we must have

$$D^2=C,\ C^3,\ C^5 \quad \text{or} \quad C^7.$$

In each case we see that C is a square and so C belongs in the principal genus. But C represents l so we must have $(l/p)=+1$. This completes the proof of (1.11). This technique of taking successive squareroots has been described by a number of authors [1], [5], [8], [10]. To complete the proof of the theorem we must show that

(1.12) $$\left(\frac{l}{p}\right)=\left(\frac{g}{p}\right)_4\left(\frac{2h}{g}\right).$$

Since l is represented primitively by the form (l, m, n) and $[l, m, n]^2=[g, 2h, 2g]$, l^2 is represented primitively by the form $(g, 2h, 2g)$. Thus there are integers x and y such that

(1.13) $$l^2=gx^2+2hxy+2gy^2, \qquad (x, y)=1.$$

Changing the signs of both x and y, if necessary, we can suppose that x is positive. Clearly x is odd. We set

(1.14) $$k=|hx+2gy|,$$

so that k is an odd positive integer. From (1.1), (1.13) and (1.14) we obtain

(1.15) $$2gl^2=k^2+px^2,$$

so that k is not divisible by p. Using (1.2), (1.10), (1.13) and (1.15), it is easy to check that

(1.16)
$$\text{G.C.D.}(x, l)=\text{G.C.D.}(x, k)=\text{G.C.D.}(x, g)=\text{G.C.D.}(k, g)=\text{G.C.D.}(k, l)=1.$$

From (1.15) we have

$$\left(\frac{2gl^2}{p}\right)_4=\left(\frac{k^2}{p}\right)_4=\left(\frac{k}{p}\right)=\left(\frac{p}{k}\right)=\left(\frac{px^2}{k}\right)=\left(\frac{2gl^2}{k}\right),$$

so that

$$\left(\frac{l}{p}\right)=\left(\frac{2}{p}\right)_4\left(\frac{g}{p}\right)_4\left(\frac{2g}{k}\right). \tag{1.16}$$

Next from (1.1) and (1.2) we obtain

$$\left(\frac{2}{p}\right)_4=\left(\frac{2g^4}{p}\right)_4=\left(\frac{g^2h^2}{p}\right)_4=\left(\frac{gh}{p}\right)=\left(\frac{h}{p}\right)=\left(\frac{p}{h}\right)=\left(\frac{2g^2}{h}\right)=\left(\frac{2}{h}\right),$$

so that (1.16) becomes

$$\left(\frac{l}{p}\right)=\left(\frac{g}{p}\right)_4\left(\frac{2}{h}\right)\left(\frac{2}{k}\right)\left(\frac{g}{k}\right). \tag{1.17}$$

Further, from (1.1) and (1.15), we get

$$k^2-1=2gl^2-(2g^2-h^2)x^2-1\equiv 2g-2+h^2x^2-1 \pmod{16},$$

so that

$$\tfrac{1}{8}(k^2-1)\equiv\tfrac{1}{4}(g-1)+\tfrac{1}{8}(h^2x^2-1) \pmod{2},$$

giving

$$\left(\frac{2}{k}\right)=\left(\frac{2}{g}\right)\left(\frac{2}{hx}\right),$$

so that (1.17) gives

$$\left(\frac{l}{p}\right)=\left(\frac{g}{p}\right)_4\left(\frac{2}{g}\right)\left(\frac{2}{x}\right)\left(\frac{g}{k}\right). \tag{1.18}$$

Finally, we have (as $g\equiv 1 \pmod 4$)

$$\begin{aligned}\left(\frac{g}{k}\right)&=\left(\frac{k}{g}\right)=\left(\frac{hx+2gy}{g}\right)=\left(\frac{hx}{g}\right)=\left(\frac{h}{g}\right)\left(\frac{g}{x}\right)\\&=\left(\frac{h}{g}\right)\left(\frac{4gl^2}{x}\right)=\left(\frac{h}{g}\right)\left(\frac{2k^2}{x}\right)=\left(\frac{h}{g}\right)\left(\frac{2}{x}\right),\end{aligned}$$

and using this in (1.18) we obtain

$$\left(\frac{l}{p}\right)=\left(\frac{g}{p}\right)_4\left(\frac{2h}{g}\right),$$

as required. This completes the proof of the theorem.

We remark that in a paper to appear elsewhere [12], the second author has shown that if $p\equiv 1 \pmod 8$ is a prime such that $h(-p)\equiv 0 \pmod 8$, then $h(-p)\equiv T+p-1 \pmod{16}$, where $T+U\sqrt{p}$ is the fundamental unit of $Q(\sqrt{p})$.

We also note that Theorem 1 answers a question of Brown [4: p. 417].

2. $D=-2p$, $p\equiv 1 \pmod 8$. In the representation (0.1) clearly u is odd and v is even. Replacing (u, v) by the representation $(3u+4v, 2u+3v)$, if necessary, we can suppose that

$$u\equiv 1 \pmod 4. \tag{2.1}$$

By a theorem of Hasse [7: p. 234] [8: p. 5], we have

$$h(-2p)\equiv 0 \pmod 8 \Leftrightarrow u\equiv 1 \pmod 8 \Leftrightarrow \left(\frac{u}{p}\right)=+1. \tag{2.2}$$

Assuming that $h(-2p)\equiv 0 \pmod 8$, in view of (2.2), the symbol $(u/p)_4$ is well-defined and independent of the choice (u, v) satisfying (0.1) and the condition $u\equiv 1 \pmod 8$. Proceeding exactly as in the proof of Theorem 1, but with I, A, B replaced by $[1, 0, 2p]$, $[2, 0, p]$, $[u, 4v, 2u]$ respectively, we obtain

$$h(-2p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{l}{p}\right)=+1,\ \left(\frac{l}{p}\right)=\left(\frac{u}{p}\right)_4,$$

which establishes the following theorem.

THEOREM 2. *Let $p\equiv 1$ (mod 8) be a prime such that $h(-2p)\equiv 0$ (mod 8). Set $p=u^2-2v^2$, where u and v are positive integers with u chosen to satisfy $u\equiv 1$ (mod 8). Then*

$$h(-2p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{u}{p}\right)_4=+1.$$

In a forthcoming paper [10], Kaplan and the second author have established a congruence modulo 16 involving $h(-2p)$ and $h(2p)$ (the narrow class number of the real quadratic field $Q(\sqrt{2p})$). Using this congruence together with Theorem 2 in the case when $p\equiv 1 \pmod 8$ is such that $h(2p)\equiv 0 \pmod 8$ (so that $p\equiv 1 \pmod{16}$) and one of the equations $x^2-2py^2=-1$ or $+2$ is solvable in integers x and y, we can obtain a necessary and sufficient condition for $h(2p)\equiv 0 \pmod{16}$.

COROLLARY. *Let $p\equiv 1$ (mod 16) be a prime such that $h(2p)\equiv 0$ (mod 8) and such that one of the equations $x^2-2py^2=-1, +2$ is solvable in integers x and y. Set $p=u^2-2v^2$, where u and v are positive integers with u chosen so that $u\equiv 1$ (mod 8). Then*

$$h(2p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{u}{p}\right)_4=+1.$$

3. $D=-2p$, $p\equiv 7 \pmod 8$. In this case it is well-known that

$$h(-2p)\equiv\begin{cases}0 \pmod 8, & \text{if} \quad p\equiv 15 \pmod{16},\\ 4 \pmod 8, & \text{if} \quad p\equiv 7 \pmod{16},\end{cases} \tag{3.1}$$

see for example [2: Cor. 7.4], [7: p. 234].

We restrict our attention to primes $p\equiv 15 \pmod{16}$. From (0.1) we deduce that $u\equiv\pm 1 \pmod 8$. Replacing the representation (u, v) by $(3u+4v, 2u+3v)$, if necessary, we can suppose that $u\equiv 1 \pmod 8$. Replacing (u, v) by $(17u+24v, 12u+17v)$, if necessary, we can further suppose that $u\equiv 1 \pmod{16}$. Again proceeding exactly as in the proof of Theorem 1, we obtain

$$h(-2p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{2}{l}\right)=+1,\ \left(\frac{2}{l}\right)=\left(\frac{v}{u}\right),$$

which establishes the following theorem.

THEOREM 3. *Let* $p\equiv 15 \pmod{16}$ *be a prime. Set* $p=u^2-2v^2$, *where* u *and* v *are positive integers with* u *chosen to satisfy* $u\equiv 1 \pmod{16}$. *Then*

$$h(-2p)\equiv 0 \pmod{16} \Leftrightarrow \left(\frac{v}{u}\right)=+1.$$

This result should be compared with the following result of the second author: if $p\equiv 15 \pmod{16}$ is prime then $h(-2p)\equiv U \pmod{16}$, where $T+U\sqrt{2p}$ is the fundamental unit of $Q(\sqrt{2p})$.

4. **Conclusion.** For $D<0$ there remains one further case when the 2-Sylow subgroup $H_2(D)$ of $H(D)$ is cyclic of order ≥ 4 (see for example [9]), namely,

$$D=-pq,\ p(\text{prime})\equiv 1 \pmod 4,\ q(\text{prime})\equiv 3 \pmod 4,\ \left(\frac{p}{q}\right)=+1. \tag{4.1}$$

In this case it is known (see for example [9: Théorème 8]) that

$$h(-pq)\equiv 0 \pmod 8 \Leftrightarrow \left(\frac{-q}{p}\right)_4=+1. \tag{4.2}$$

It would be interesting to obtain an explicit necessary and sufficient condition for $h(-pq)\equiv 0 \pmod{16}$ in this case too, but since (4.2) already involves the Dirichlet symbol $(-q/p)_4$ this may be difficult.

The authors would like to thank Pierre Kaplan of the University of Nancy, France for a helpful comment in connection with the proof of Theorem 1. Kaplan [10] has obtained various criteria for the existence of cycles of order 16 in the class group of certain quadratic fields.

REFERENCES

1. Helmut Bauer, *Zur Berechnung der 2-Klassenzahl der quadratische Zahlkörper mit genau zwei verschiedenen Diskriminantenprimteilern*, J. Reine Angew. Math. **248** (1971), 42–46.

2. Bruce C. Berndt, *Classical theorems on quadratic residues*, L'Enseignement Math. **22** (1976), 261–304.

3. Ezra Brown, *The class number of* $Q(\sqrt{-p})$, *for* $p \equiv 1 \pmod 8$ *a prime*, Proc. Amer. Math. Soc. **31** (1972), 381–383.

4. Ezra Brown, The power of 2 dividing the class-number of a binary quadratic discriminant, J. Number Theory **5** (1973), 413–419.

5. Ezra Brown, *Class numbers of quadratic fields*, Symp. Mat. **15** (1975), 403–411.

6. Helmut Hasse, *Über die Klassenzahl des Körpers* $P(\sqrt{-p})$ *mit einer Primzahl* $p \equiv 1 \pmod{2^3}$, Aequationes Math. **3** (1969), 165–169.

7. Helmut Hasse, Über die Klassenzahl des Körpers $P(\sqrt{-2p})$ mit einer Primzahl $p \neq 2$, J. Number Theory 1 (1969), 231–234.

8. Helmut Hasse, Über die Teilbarkeit durch 2^3 der Klassenzahl imaginär-quadratischer Zahlkörper mit genau zwei verschiedenen Diskriminantenprimteilern, J. Reine Angew. Math. 241 (1970), 1–6.

9. Pierre Kaplan, Divisibilité par 8 du nombre des classes des corps quadratiques dont le 2-groupe des classes est cyclique, et réciprocité biquadratique, J. Math. Soc. Japan 25 (1973), 596–608.

10. Pierre Kaplan, Cycles d'ordre au moins 16 dans le 2-groupe des classes d'idéaux de certains corps quadratiques, Bull. Soc. Math. France Mém. No. 49–50 (1977), 113–124.

11. Pierre Kaplan and Kenneth S. Williams, On the class numbers of $Q(\sqrt{\pm 2p})$ modulo 16, for $p \equiv 1 \pmod 8$ a prime, Acta Arith. (to appear).

12. Kenneth S. Williams, On the class number of $Q(\sqrt{-p})$ modulo 16, for $p \equiv 1 \pmod 8$ a prime, Acta Arith. (to appear).

ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA, U.S.A.
82581

CARLETON UNIVERSITY
OTTAWA, ONTARIO, CANADA
K1S 5B6

Canad. Math. Bull. Vol. **25** (2), 1982

ON DUNFORD–PETTIS OPERATORS

BY
ELIAS SAAB

ABSTRACT. Let X be a complemented subspace of a Banach lattice E. It is shown that if every Dunford–Pettis operator from $L_1[0, 1]$ into X is Pettis-representable then X has the Radon–Nikodym property.

In [1] Bourgain showed that if every Dunford–Pettis operator from $L_1[0, 1]$ to a Banach space X is Bochner-representable then X has the Radon–Nikodym property. In this paper we show that if X is complemented in a Banach lattice E and if every Dunford–Pettis operator from $L_1[0, 1]$ into X is Pettis-representable then X has the Radon–Nikodym property.

All the notions used in this paper and not defined can be found in ([2], [4], [6]). Let E be a Banach space and let T be an operator from $L_1[0, 1]$ into E.

DEFINITION 1. (i) The operator T is said to be Dunford–Pettis if the set $\{T(1_A);\ A$ is a measurable subset of $[0, 1]\}$ is relatively compact in E.

(ii) The operator T is said to be Bochner- (resp. Pettis) representable if there exists $g:[0, 1]\to E$ Bochner integrable and essentially bounded (resp. Pettis integrable and scalarly essentially bounded) such that for every f in $L_1[0, 1]$, $T(f) = \text{Bochner-}\int_0^1 fg\,d\lambda$ (resp., $T(f) = \text{Pettis-}\int_0^1 fg\,d\lambda$).

It is well known that the Dunford–Pettis operators are precisely those which map weakly convergent sequences into norm convergent sequences, it is also known that a Pettis representable operator is Dunford–Pettis and that a Dunford–Pettis operator is not in general Pettis-representable.

Bourgain showed in [1] that a Banach space E has the Radon–Nikodym property if and only if every Dunford–Pettis operator is Bochner-representable; he also constructed an operator $T: L_1[0, 1]\to c_0$ such that T is Dunford–Pettis but T is not even Pettis-representable in l_∞.

Because this operator is useful in the sequel we are going to describe it quickly. First construct a sequence $(A_n)_{n\geq 1}$ of measurable subsets of $[0, 1]$ such

---

Received by the editors May 12, 1980 in revised form, January 20, 1981.
AMS (MOS) Subject Classifications. 46G10, 46B22.
Key words and phrases. Dunford–Pettis operators, Radon–Nikodym property.

that

(i) $\lim_n \lambda(A_n) = 0$

(ii) $\{1_{A_n}\}_{n\geq 1}$ is dense in $\{0, 1\}^{[0,1]}$ endowed with the product topology.

It is clear that $T: L_1[0, 1] \to c_0$ defined by $T(f) = (\int_{A_n} f\,d\lambda)_{n\geq 1}$ is Dunford–Pettis and by using a non measurable cluster point of the sequence $\{1_{A_n}\}_{n\geq 1}$ one can show that T is not Pettis-representable in l_∞.

DEFINITION 2. A Banach space E has the separable complementation property if every separable subspace Y of E is contained in a complemented separable subspace Z of E.

LEMMA 3. *Let E be a Banach space having the separable complementation property and let F be a subspace of E. If $T: L_1[0, 1] \to F$ is Pettis-representable in F, then it is Bochner-representable in F.*

Proof. Let Y = the closure of $T(L_1[0, 1])$. Let Z be a separable subspace of E containing Y and complemented in E by a projection $V: E \to Z$. Let g be the Pettis derivative of T. The map $t \to V(g(t))$ from $[0, 1] \to Z$ is strongly measurable and essentially bounded and hence is Bochner integrable. Therefore for every f in $L_1[0, 1]$

$$T(f) = V(T(f)) = V\left(\text{Pettis-}\int_0^1 fg\,d\lambda\right)$$
$$= \text{Pettis-}\int_0^1 fV(g)\,d\lambda = \text{Bochner-}\int_0^1 fV(g(t))\,d\lambda.$$

This implies that the map $t \to V(g(t)$ takes its values λ-almost everywhere in F and it is the Bochner derivative of T in F.

PROPOSITION 4. *Let E be a Banach space such that every Dunford–Pettis operator from $L_1[0, 1]$ into E is Pettis-representable, then E does not contain a subspace isomorphic to c_0.*

Proof. Suppose that c_0 is isomorphic to a subspace of E, let $S: c_0 \to E$ be this isomorphism, the double adjoint S^{**} of S embeds l_∞ in E^{**}. Let U be a projection from E^{**} to $S^{**}(l_\infty)$ and consider the following diagram:

$$\begin{tikzcd} c_0 \arrow[r, "S"] & E \arrow[r, "Q"] & E^{**} \arrow[r, "U"] & S^{**}(l_\infty). \end{tikzcd}$$

Let $T: L_1[0, 1] \to c_0$ the Dunford–Pettis operator constructed by Bourgain. By hypothesis SoT is Pettis-representable in E and hence by the above diagram T will be Pettis-representable in l_∞, a contradiction that finishes the proof.

COROLLARY 5. *If a Banach space has the weak-Radon Nikodym property then E does not contain any isomorphic copy of c_0.*

The only known proofs of the above Corollary 5 ([4], [5]) rely heavily either on a deep result of Fremlin [3] or of Sierpinski [7].

The following fact was shown in [4]. We are going to give a slightly different proof of it.

PROPOSITION 6. *Let E be an order continuous Banach lattice. Then E has the separable complementation property.*

Proof. Let Y be a separable subspace of E. Let $(x_n)_{n\geq 1}$ be a dense subset of the positive unit ball of Y. Consider

$$u=\sum_{n=1}^{\infty}\frac{x_n}{2^n}.$$

Let F be the closed ideal generated by u, i.e.

$$F=\overline{\bigcup_{n=1}^{\infty}[-nu, nu]}.$$

It is clear that F contains Y. Since E is order continuous, the space F is weakly compactly generated and F is complemented in E. Choose Z a separable subspace of F which is complemented in F and containing Y. It is clear that Z is complemented in E.

Combining the results of [1] and ([6], p. 36) with Lemma 3, Proposition 4, and Proposition 6 we get

THEOREM 7. *Let X be a complemented subspace of a Banach lattice E. If every Dunford–Pettis operator from L_1 to X is Pettis-representable then X has the Radon–Nikodym property.*

REFERENCES

1. J. Bourgain, Dunford-Pettis Operators and the Radon–Nikodym property, (preprint).

2. J. Diestel and J. J. Uhl, Jr., *Vector measures*, Mathematical Survey No. 15, American Mathematical Society, Providence, 1977.

3. D. H. Fremlin, *Pointwise compact subsets of measurable functions*, Manuscripts Math. **15**, 219–242 (1975).

4. N. Ghoussoub and E. Saab, *On the weak Radon–Nikodym property*, Proc. Amer. Math. Soc. **81** (1981), 81–84.

5. L. Janicka, Wlasnosci Typu Radona–Nikodyma dla Przestrzeni Banacha, Thesis, 1978, Wroclaw.

6. J. Lindenstrauss and L. Tzafriri, Classical Banach spaces II, Springer-Verlag (1979).

7. W. Sierpinski, *Fonctions additives non completement additives et fonctions non mesurables*. Fund. Math. **30** (96–99) 1938.

THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C. V6T 1Y4
CANADA

Current address:
THE UNIVERSITY OF MISSOURI
COLUMBIA, MO 65211

Canad. Math. Bull. Vol. **25** (2), 1982

UNIFORMLY LIPSCHITZIAN SEMIGROUPS IN HILBERT SPACE

BY
DAVID J. DOWNING AND WILLIAM O. RAY[1]

ABSTRACT. Let K be a closed, bounded, convex, nonempty subset of a Hilbert Space $\mathscr{H}$. It is shown that if $\mathscr{T}$ is a left reversible, uniformly k-lipschitzian semigroup of mappings of K into itself, with $k<\sqrt{2}$, then $\mathscr{T}$ has a common fixed point in K.

1. **Introduction.** Let $\mathscr{T}=\{T_\alpha\}_{\alpha\in A}$ be a semigroup of mappings of a metric space (M, d) into itself. Such a semigroup is said to have a *common fixed point* if there exists $x_0\in M$ with $T_\alpha(x_0)=x_0$ for all $\alpha\in A$; $\mathscr{T}$ is said to be *uniformly k-lipschitzian* semigroup if, for each $x, y\in M$ and $\alpha\in A$,

$$d(T_\alpha(x), T_\alpha(y))\le k\, d(x, y).$$

Uniformly k-lipschitzian semigroups were introduced (in a slightly more general form) by K. Goebel, W. A. Kirk, and R. L. Thele in [2], and they also assumed that the semigroup $\mathscr{T}$ was *left reversible* (i.e., every two right ideals in $\mathscr{T}$ have non-empty intersection). This latter is automatically fulfilled if, for example $\mathscr{T}$ is commutative, and in particular if $\mathscr{T}=\{T_s\}_{s\in[0,\infty)}$. The basic result of [2] asserts that if E is a uniformly convex Banach space then there is a $k_0>1$ such that, whenever $K\subseteq E$ is a closed, bounded, convex set and $\mathscr{T}$ is a left reversible uniformly k-lipschitzian semigroup of mappings from K into K with $k<k_0$, then $\mathscr{T}$ has a common fixed point in K. Precisely how large k_0 may be taken to be remains, even in Hilbert space, an open question; the estimate provided for Hilbert space in [2] was $\sqrt{5}/2$, with an upper bound of 2. In the special case where $\mathscr{T}$ consists of iterates of a single mapping $T: K\to K$, T is said to be uniformly k-lipschitzian mapping. These mappings were first studied by K. Goebel and W. A. Kirk in [1]. In [4], E. Lifschitz proved, using a technique different from the one we employ below, that in Hilbert space a uniformly k-lipschitzian mapping with $k<\sqrt{2}$ has a fixed point. Our main purpose in this note, accomplished in Section 2, is to show that the estimate of $\sqrt{2}$ is valid under the more general semigroup assumptions.

Received by the editors May 12, 1980 and, in revised form, November 3, 1980.
AMS(MOS) subject classification: 47H10.
Key words: common fixed points, uniformly k-lipschitzian semigroups.

[1] Research carried out in part while at Iowa State University, Ames, Iowa.

2. **Uniformly k-Lipschitzian semigroups.** The main result of this note may be stated as follows:

THEOREM 1. *Let $\mathcal{H}$ be a Hilbert space and let K be a nonempty, closed, convex, bounded subset of $\mathcal{H}$. Let $\mathcal{T}=\{T_\alpha\}_{\alpha\in A}$ be a left reversible semigroup of mappings: $T_\alpha: K\to K$ for each $\alpha\in A$. If $\mathcal{T}$ is uniformly k-lipschitzian with $k<\sqrt{2}$, then there exists $x_0\in K$ with $T_\alpha(x_0)=x_0$ for all $\alpha\in A$.*

The basic idea of our proof is the same as the proof of Theorem 2.1 in [2], and we include many of the details only for the sake of completeness; our result requires, however, somewhat more refined bounds on the quantity $d(x)$ defined below. These bounds, in turn, are motivated by a result of N. Routledge [5] (c.f., [3], page 192) which asserts that, in Hilbert space, the diameter of a set is equal to $\sqrt{2}$ times the optimal Chebyshev radius of the set.

Proof of Theorem 1. We may assume $k>1$. For each $\alpha\in A$, let $\mathcal{T}_\alpha=\{T_\alpha\circ T: T\in\mathcal{T}\}$ and for each $x\in K$, let $\mathcal{T}_\alpha(x)=\{T(x): T\in\mathcal{T}_\alpha\}$. In addition set $d(x)=\inf_{\alpha\in A}\{\sup\|x-Tx\|: T\in\mathcal{T}_\alpha\}$. It will suffice to show $d(x_0)=0$ for some $x_0\in K$. For suppose this is the case; since $\mathcal{T}$ is left reversible, the family $\{T_\alpha\}_{\alpha\in A}$ forms a directed set under the relation:

$$\mathcal{T}_\alpha\geq\mathcal{T}_\beta \quad\text{if and only if}\quad \mathcal{T}_\alpha\subseteq\mathcal{T}_\beta. \tag{1}$$

Now if $x_\alpha\in\mathcal{T}_\alpha(x_0)$ for each $\alpha\in A$, the fact that $d(x_0)=0$ yields that the net $\{x_\alpha\}_{\alpha\in A}$ converges to x_0; and thus, if $T\in\mathcal{T}$, $\{Tx_\alpha\}_{\alpha\in A}$ converges to Tx_0. But for $T\in\mathcal{T}$, $\{T\mathcal{T}_\alpha(x_0)\}_{\alpha\in A}$ is a subset of $\{\mathcal{T}_\alpha(x_0)\}$ and $Tx_\alpha\in T\mathcal{T}_\alpha(x_0)$ for all $\alpha\in A$. This implies that the net $\{Tx_\alpha\}_{\alpha\in A}$ converges to x_0, whence $Tx_0=x_0$ for all $T\in\mathcal{T}$.

Now to see $d(x_0)=0$ for some $x_0\in K$, fix $x\in K$. Let $R(x)=\{r>0: \mathcal{T}_\alpha(x)\subseteq B(y;r)$ for some $\alpha\in A$ and $y\in K\}$ and let $r_0(x)=r_0=\inf R(x)$. Note that if $r<r_0(x)$ and $z\in K$, then for all $\alpha\in A$, there exists $T\in\mathcal{T}_\alpha$ with

$$\|z-Tx\|>r. \tag{2}$$

Let $\varepsilon>0$ and set

$$D(r_0,\alpha,\varepsilon)=\bigcap_{T\in\mathcal{T}_\alpha} B(Tx; r_0+\varepsilon)\cap K.$$

Clearly for each $\varepsilon>0$, there exists $\alpha\in A$ with $D(r_0,\alpha,\varepsilon)\neq\emptyset$. Also, for fixed ε, the family $\{D(r_0,\alpha,\varepsilon)\}_{\alpha\in A}$ is an increasing net when directed as in (1). Thus if $C_\varepsilon=\bigcup_{\alpha\in A} D(r_0,\alpha,\varepsilon)$, C_ε is nonempty and convex. It then follows that if $C=\bigcap_{\varepsilon>0}(clC_\varepsilon\cap K)$, C is also nonempty. Let $g(x)\in C$. We may assume $r_0>0$; for if $r_0=0$, then for each $\varepsilon>0$ there exists $\alpha\in A$ with $\|g(x)-Tx\|\leq\varepsilon$ for all $T\in\mathcal{T}_\alpha$. Thus for each $T\in\mathcal{T}_\alpha$,

$$\begin{aligned}\|g(x)-Tg(x)\|&\leq\|g(x)-T^2x\|+\|T^2x-Tg(x)\|\\&\leq\varepsilon+k\,\|Tx-g(x)\|\\&\leq\varepsilon(1+k)\end{aligned}$$

and so $d(g(x))=0$. Fix $\varepsilon>0$, $\varepsilon<\min\{r_0/2,\ d(g(x))/2\}$, and $\lambda\in[0,1]$. Choose $\alpha\in A$ so that

$$\|g(x)-T_\alpha(g(x))\|\geq d(g(x))+\varepsilon, \tag{3}$$

and choose $\beta\in A$ with

$$\|g(x)-Tx\|\leq r_0+\varepsilon \quad \text{for all} \quad T\in\mathcal{T}_\beta.$$

Now, since $\mathcal{T}$ is a semigroup, $T_\alpha\circ T_\beta=T_\gamma$ for some $\gamma\in A$. Let $\mu\in A$ so that $T_\mu\in\mathcal{T}_\gamma\cap\mathcal{T}_\beta$; then $\mathcal{T}_\mu\subseteq\mathcal{T}_\gamma\cap\mathcal{T}_\beta$. If $T\in\mathcal{T}_\mu$, there exists $\tilde{T}\in\mathcal{T}_\beta$ with $T=T_\alpha\circ\tilde{T}$. This yields for $T\in\mathcal{T}_\mu$

$$\begin{aligned}\|T_\alpha(g(x))-Tx\|&=\|T_\alpha(g(x))-T_\alpha\circ\tilde{T}x\|\\&\leq k\,\|g(x)-\tilde{T}x\|\leq k(r_0+\varepsilon),\end{aligned} \tag{4}$$

and since $\mathcal{T}_\mu\subseteq\mathcal{T}_\beta$,

$$\|g(x)-Tx\|\leq r_0+\varepsilon \tag{5}$$

for all $T\in\mathcal{T}_\mu$. Finally, by (2), we may choose $T_0\in\mathcal{T}_\mu$ with

$$\|(1-\lambda)T_\alpha(g(x))+\lambda g(x)-T_0(x)\|\geq r_0-\varepsilon. \tag{6}$$

Set $u=g(x)-T_0x$, $v=T_\alpha(g(x))-T_0x$, so $u-v=g(x)-T_\alpha(g(x))$. By (6), $\|\lambda u+(1-\lambda)v\|\geq r_0-\varepsilon$ and so by (4) and (5) we have

$$(r_0-\varepsilon)^2\leq\|\lambda u+(1-\lambda)v\|^2\leq\lambda^2(r_0+\varepsilon)^2+2\lambda(1-\lambda)\langle u,v\rangle+k^2(1-\lambda)^2(r_0+\varepsilon)^2$$

thus

$$(r_0-\varepsilon)^2-\lambda^2(r_0+\varepsilon)^2-k^2(1-\lambda)^2(r_0+\varepsilon)^2\leq2\lambda(1-\lambda)\langle u,v\rangle$$

or

$$-2\langle u,v\rangle\leq\frac{-(r_0-\varepsilon)^2+\lambda^2(r_0+\varepsilon)^2+k^2(1-\lambda)^2(r_0+\varepsilon)^2}{\lambda(1-\lambda)}.$$

Using this, we obtain

$$\begin{aligned}\|u-v\|^2&\leq(r_0+\varepsilon)^2-2\langle u,v\rangle+k^2(r_0+\varepsilon)^2\\&\leq(r_0+\varepsilon)^2+\frac{-(r_0+\varepsilon)^2+\lambda^2(r_0+\varepsilon)^2+k^2(1-\lambda)^2(r_0+\varepsilon)^2}{\lambda(1-\lambda)}+k^2(r_0+\varepsilon)^2\\&=\frac{\begin{aligned}\lambda(1-\lambda)(r_0+\varepsilon)^2-(r_0-\varepsilon)^2+\lambda^2(r_0+\varepsilon)^2+k^2(1-\lambda)^2(r_0+\varepsilon)^2\\+\lambda(1-\lambda)k^2(r_0+\varepsilon)^2\end{aligned}}{\lambda(1-\lambda)}\end{aligned} \tag{7}$$

By (3),

$$(d(g(x))+\varepsilon)^2\leq\|u-v\|^2$$

Combining this inequality with (7) above and taking the limit as $\varepsilon \to 0$,

$$\begin{aligned} d(g(x))^2 &\le \frac{\lambda(1-\lambda)r_0^2 - r_0^2 + \lambda^2 r_0^2 + k^2(1-\lambda)^2 r_0^2 + \lambda(1-\lambda)k^2 r_0^2}{\lambda(1-\lambda)} \\ &= \frac{\lambda(1-\lambda)r_0^2 - (1-\lambda^2)r_0^2 + k^2(1-\lambda)^2 r_0^2 + \lambda(1-\lambda)k^2 r_0^2}{\lambda(1-\lambda)} \\ &= \frac{\lambda r_0^2 - (1+\lambda)r_0^2 + k^2(1-\lambda)r_0^2 + \lambda k^2 r_0^2}{\lambda}. \end{aligned}$$

Letting $\lambda \to 1$,

$$d(g(x))^2 \le (k^2-1)r_0^2 \quad \text{or} \quad d(g(x)) \le \sqrt{(k^2-1)}\,r_0$$

It is clear that $r_0(x) \le d(x)$ and that

$$\|g(x)-x\| \le r_0(x) + d(x) \le 2d(x).$$

Thus, for some $\xi < 1$ ($\xi = \sqrt{(k^2-1)}$) and for each $x \in K$, we have shown that there exists $g(x) \in K$ with

$$d(g(x)) \le \xi d(x), \qquad \|g(x)-x\| \le 2d(x).$$

Define a sequence $\{x_n\}$ in K by fixing $x_0 \in K$ and letting $x_{n+1} = g(x_n)$ for $n = 0, 1, 2, \ldots$. If $r_0(x_n)$ or $d(x_n) = 0$ for any n, we are done. Otherwise note

$$\|x_{n+1} - x_n\| \le 2d(x_n) \le 2\xi^n\, d(x_0),$$

so that $\{x_n\}$ is a Cauchy sequence. Therefore $x_n \to z \in K$ as $n \to \infty$. Let $\{\varepsilon_m\}$ be a sequence of positive numbers with $\varepsilon_m \to 0$, and for each n, choose $\alpha \in A$ so that

$$\|x_n - Tx_n\| \le d(x_n) + \varepsilon_n \quad \text{for all} \quad T \in \mathcal{T}_{\alpha_n}.$$

Then for $T \in \mathcal{T}_{\alpha_n}$,

$$\begin{aligned} \|z - Tz\| &\le \|z - x_n\| + \|x_n - Tx_n\| + \|Tx_n - Tz\| \\ &\le (1-k)\,\|z - x_n\| + d(x_n) + \varepsilon_n. \end{aligned}$$

This quantity can be made arbitrarily small, hence $d(z) = 0$.

REFERENCES

1. K. Goebel and W. A. Kirk, *A fixed point theorem for transformations whose iterates have uniform Lipschitz constant*, Studia Math., **47** (1973), 135–140.

2. K. Goebel, W. A. Kirk, and R. L. Thele, *Uniformly lipschitzian families of transformations in Banach spaces*, Can. J. Math., **26** (1974), 1245–1256.

3. R. B. Holmes, *A Course on Optimization and Best Approximation*, Lecture Notes No. 257, Springer-Verlag, Berlin, Heidelberg, New York, 1972.

4. E. A. Lifschitz, *Fixed point theorems for operators in strongly convex spaces*, Voronež Gos. Univ. Trudy Math. Fak., **16** (1975), 23–28. (Russian)

5. N. Routledge, *A result in Hilbert space*, Quarterly J. Math., **3** (1952), 12–18.

6. R. H. Martin Jr., *Nonlinear Operators and Differential Equations in Banach Spaces*, Wiley-Interscience, New York, London, Sydney, Toronto, 1976.

Department of Mathematical Sciences
Oakland University
Rochester, Michigan 48063

Department of Mathematics
The University of Oklahoma
Norman, Oklahoma 73019

Canad. Math. Bull. Vol. **25** (2), 1982

A PRESENTATION FOR A GROUP OF INTEGER MATRICES

BY

N. SMYTHE

ABSTRACT. Let Λ_n be the kernel of the map from $GL(n, \mathbb{Z})$ to $GL(n, \mathbb{Z}_2)$ induced by the quotient map $\mathbb{Z} \to \mathbb{Z}_2$. We give a presentation for Λ_n.

The kernel of the map from $GL(n, \mathbb{Z})$ to $GL(n, \mathbb{Z}_p)$ is of some interest in K-theory. We give here a presentation for this group in the case $p=2$; we denote the group by Λ_n. Since a general presentation for $GL(n, \mathbb{Z})$ and coset representatives for $GL(n, \mathbb{Z}_2)$ are known (see Coxeter and Moser [1]), a presentation can of course be obtained by using the Reidemeister–Schreier algorithm for any particular value of n. However it is not clear how to do this for general n; and in any case the presentation derived in this way is enormous and essentially useless for deriving properties of the group.

1. **Generators and canonical form.** The group Λ_n consists of $n \times n$ integer matrices of determinant ± 1 with odd entries on the diagonal, even entries off diagonal. Let Γ_n be the subgroup of Λ_n consisting of matrices whose diagonal entries are congruent to 1 modulo 4; also let A_i be the diagonal matrix with entry -1 in the ith position, $+1$ elsewhere, and let P_n be the subgroup of Λ_n generated by $A_1, A_2, \ldots, A_n$. Clearly Λ_n is generated by $\Gamma_n P_n$, the subgroup Γ_n is normal in Λ_n, P_n is an elementary 2-group of rank n, and Λ_n is a semi-direct product of Γ_n and P_n. Finally it should be noted that $A_1 A_2 \cdots A_n = -I$ is the central element of $GL(n, \mathbb{Z})$, and the group $\Lambda_n^+ = \Lambda_n \cap SL(n, \mathbb{Z})$ is the semi-direct product of Γ_n and the subgroup of P_n generated by A_1A_2, $A_1A_3, \ldots, A_1A_n$.

Let B_{ij} denote the matrix which is the identity except for an entry 2 in the (i, j) position. Then Γ_n is generated by $\{B_{ij} : 1 \le i, j \le n, i \ne j\}$. Moreover we can write each element of Γ_n as a word in these generators in a canonical way:

THEOREM 1. *Each element of Γ_n has a unique expression as a product*

$$(\beta_{12}\beta_{13} \cdots \beta_{1n})(\beta_{23}\beta_{24} \cdots \beta_{2n}) \cdots (\beta_{n-1n})(\delta_{n-1n} \cdots \delta_{12})$$

where β_{ij} is a reduced word in the free group $\langle B_{ij}, B_{ji}\rangle$ of degree zero in the generator B_{ij}, $i<j$, and δ_{ij} is a power of B_{ij}, $i<j$.

Received by the editors, June 11, 1980 and in revised form, January 22, 1981.
1980 AMS Classifications 20FO5, 2OH25.

Proof. Let M be any matrix in Γ_n. Left multiplication of M by B_{12} adds twice the second row of M to the first, and left multiplication by B_{21} adds twice the first row to the second. Hence by left multiplication by elements of $\langle B_{12}, B_{21}\rangle$ we can successively reduce the absolute values of the first two entries in the first column of M. Since the first entry remains odd, there is an element $\beta_{12} \in \langle B_{12}, B_{21}\rangle$ such that $\beta_{12}^{-1}M$ has a zero in the (2, 1) position; furthermore, we may assume that β_{12} has degree zero in B_{12}. Proceeding down the first column, we produce elements $\beta_{1j} \in \langle B_{ij}, B_{j1}\rangle$ $j \leq n$ of degree zero in B_{1j}, so that

$$M' = \beta_{1n}^{-1}\beta_{1n-1}^{-1} \cdots \beta_{12}^{-1}M$$

has first column consisting of zeros except in the leading entry, which then must be 1. By right multiplication by suitable powers of $B_{1n}, \ldots, B_{12}$ the first row of M' can then be cleared, giving

$$M = (\beta_{12}\beta_{13} \cdots \beta_{1n})M''(\delta_{1n} \cdots \delta_{12})$$

with

$$M'' = \begin{pmatrix} 1 & 0 \\ 0 & M^* \end{pmatrix},$$

M^* an $(n-1)\times(n-1)$ matrix in Γ_{n-1}.

We may assume by induction that

$$M'' = (\beta_{23} \cdots \beta_{2n}) \cdots \beta_{n-1n}\delta_{n-1n} \cdots \delta_{23}$$

uniquely.

Suppose that M can also be written

$$M = (\gamma_{12}\gamma_{13} \cdots \gamma_{1n})M'''(\varepsilon_{1n} \cdots \varepsilon_{12})$$

where $\gamma_{ij} \varepsilon \langle B_{1j}, B_{j1}\rangle$, γ_{1j} of degree zero in B_{1j}, ε_{1j} being a power of B_{1j} and $M''' \in \langle B_{ij}, B_{ji} : 2 \leq i < j \leq n\rangle$.

Then

$$\gamma_{1n}^{-1} \cdots \gamma_{12}^{-1}\beta_{12} \cdots \beta_{1n} = M'''\varepsilon_{1n} \cdots \varepsilon_{12}\delta_{12}^{-1} \cdots \delta_{1n}^{-1}(M'')^{-1}$$

The right hand side is a matrix of the form

$$\begin{pmatrix} 1 & v \\ 0 & M^{**} \end{pmatrix}$$

where v is a $1\times(n-1)$ row of integers. From the form of the left hand side it is then immediate that

$$\gamma_{12}^{-1}\beta_{12} = \left(\begin{array}{cc:c} a & b & \\ & & 0 \\ 0 & c & \\ \hdashline & 0 & I \end{array}\right)$$

The only matrices in $\langle B_{12}, B_{21}\rangle$ of this form are powers of B_{12}, and since by assumption the degree of B_{12} in the product is zero (the group $\langle B_{12}, B_{21}\rangle$ is well-known to be free), we have $\beta_{12} = \gamma_{12}$. Similarly $\beta_{1j} = \gamma_{1j}$ for each j.

Finally, applying the same argument to the resulting equation

$$\varepsilon_{1n} \cdots \varepsilon_{12}\delta_{12}^{-1} \cdots \delta_{1n}^{-1} = (M''')^{-1}M = \begin{pmatrix} 1 & 0 \\ 0 & \bar{M} \end{pmatrix}$$

we see that $\varepsilon_{1j} = \delta_{1j}$ and $M''' = M''$. The product form for M is therefore unique.

2. **Relations for** Γ_n. It is straightforward to verify that the following relations hold in Γ_n:

(a) $B_{ij} \leftrightarrow B_{ik}$
(b) $B_{ji} \leftrightarrow B_{ki}$
(c) $[B_{ij}, B_{jk}] = B_{ik}^2$
(d) $[B_{ij}B_{kj}, B_{ji}B_{jk}^{-1}] = (B_{ik}B_{ki}^{-1})^2, [B_{ji}B_{jk}, B_{ij}B_{kj}^{-1}] = (B_{ik}B_{ki}^{-1})^2$
(e) $B_{ij} \leftrightarrow B_{kl}$

whenever i, j, k, l are distinct indices. Here "$X \leftrightarrow Y$" means X commutes with Y, and $[X, Y] = XYX^{-1}Y^{-1}$; in (d) one may assume $i < j < k$.

THEOREM 2. *The above form a complete set of relations for* Γ_n.

Proof. That the relations (a), (b), (c), (d) present the group Γ_3 can be verified using the Reidemeister–Schreier algorithm, using for example the presentation for $GL(3, \mathbb{Z})$ given in Coxeter and Moser [1]. The calculation is tedious, but can be done by hand; it has been verified by computer.

(A direct proof that these relations suffice to enable the reduction of any word in Γ_3 to canonical form should be possible, but my attempts at this lead to unsolved problems in number theory involving primes having 2 or -2 as a primitive root; partial results which are obtainable however certainly help greatly in the hand reduction of the enormous Reidemeister–Schreier presentation to the given one.)

We shall assume the inductive hypothesis that Γ_n has the above presentation for $3 \leq n < N$. In order to prove the result for Γ_N, it suffices to show that every word in the generators B_{ij} can be reduced to the canonical form by means of the relations. Let G_N be the group actually presented by the relations; for any set of integers $\{i_1, i_2, \ldots, i_r\}$ with $i_1 < i_2 < \cdots < i_r$ let

$\langle i_1, i_2, \ldots, i_r\rangle$ denote the subgroup of G_N generated by all $B_{i_j i_k}$;

$\Delta(i_1, i_2, \ldots, i_r)$ denote the subgroup of G_N generated by those $B_{i_j i_k}$ with $i_j < i_k$;

$[i_1, i_2]$ denote the subgroup of $\langle i_1, i_2\rangle$ consisting of words of degree zero in $B_{i_1 i_2}$; (Note $\langle i_1, i_2\rangle$ is free since its image in Γ_N is free).

Then $G_N = \Gamma_N$ if

$$G_N = [1, 2][1, 3] \cdots [1, N][2, 3] \cdots [N-1, N]\Delta(N-1, N) \cdots \Delta(1, 2)$$

The right hand side is

$$[1, 2][1, 3] \cdots [1, N]\langle 2, 3, \ldots, N\rangle\Delta(1, N)\Delta(1, N-1) \cdots \Delta(1, 2) \tag{1}$$

by the induction assumption. (From the diagram

$$\begin{tikzcd}
G_{N-1} \arrow[r] \arrow[d, equal] & G_N \arrow[r] & \Gamma_N \arrow[d, equal] \\
\Gamma_{N-1} \arrow[rr, "\text{embedding}"] & & \Gamma_N
\end{tikzcd}$$

we may identify $\langle 2, 3, \ldots, N\rangle$ in G_N with G_{N-1}.)

Since the theorem is true for $N=3$, we note that

$$\begin{aligned}\langle i, j, k\rangle &= [1, j][i, k][j, k]\Delta(i, j, k) \quad \text{for all} \quad i<j<k \\ &= [i, k][i, j][j, k]\Delta(i, j, k)\end{aligned}$$

(a) Canonical form for $\Delta(1, 2, \ldots, \mathrm{N})$.

Let Δ_i denote the subgroup of G_N generated by $\{B_{ij} : j>i\}$. Then Δ_i is an abelian group (in fact free abelian of rank $N-i$) and is normal in $\Delta(i, i+1, \ldots, N)$ since

$$\begin{cases} B_{kl}B_{ij}B_{kl}^{-1} = B_{ij} & \text{if} \quad i, j, k, l \text{ are all distinct} \\ & \text{or if} \quad i=k \quad \text{or} \quad j=l \\ B_{jk}B_{ij}B_{jk}^{-1} = B_{ik}^{-2}B_{ij} & \text{if} \quad i<j<k \end{cases}$$

Hence

$$\begin{aligned}\Delta(1, 2, \ldots, N) &= \Delta_{N-1}\Delta_{N-2}\cdots\Delta_1 \\ &= \Delta(N-1, N)\cdots\Delta(1, N)\Delta(1, N-1)\cdots\Delta(1, 2).\end{aligned}$$

(b) Again from these relations we find that the subgroup $\langle B_{ij}, B_{kj}\rangle$ is normal in $\langle B_{ij}, B_{kl}, B_{ik}, B_{ki}\rangle$, and similarly $\langle B_{ij}, B_{ik}\rangle$ is normal in $\langle B_{ij}, B_{ik}, B_{jk}, B_{kj}\rangle$; and further

(i) $\Delta(i, j)[i, k] \subset [i, k]\Delta(k, j)\Delta(i, j)$ for $i<k<j$
(ii) $\Delta(i, j)[j, k] \subset [j, k]\Delta(i, k)\Delta(i, j)$ for $i<j<k$
(iii) $\Delta(i, j)[k, i] \subset [k, i]\Delta(k, j)\Delta(i, j)$ for $k<i<j$
(iv) $\Delta(i, j)[k, j] \subset [k, j]\Delta(i, k)\Delta(i, j)$ for $i<k<j$
(v) $\Delta(i, j)[i, k] \subset [i, k][j, k]\Delta(i, j)$ for $i<j<k$

From these we can quickly deduce

(vi) $\Delta_1 \cdot \langle 2, 3, \ldots, N\rangle \subset \langle 2, 3, \ldots, N\rangle \cdot \Delta_1$
(vii) $\Delta_1 \cdot [1, k] \subset [1, k]\langle 2, 3, \ldots, n\rangle \cdot \Delta_1$

(C) Canonical form for G_N.

Now let w be any word in the generators of G_N of length k say. By induction on k, we may assume that the initial section of w is already in canonical form, so

$$w \in [1, 2]\cdots[1, N]\langle 2, 3, \ldots, N\rangle\Delta_1 \cdot B_{ji}^{\nu} \quad \text{for some } i, j, \nu.$$

By (a) above, if $j<i$ then $\Delta_1 \cdot B_{ji}^{\nu}$ can be put in canonical form, so that w lies in the group (1).

Thus we may assume that $i<j$; then $B_{ji}^{v}\in[i,j]$ and using (b) we see that

$$w\in[1,2][1,3]\cdots[1,N]\langle 2,3,\ldots,N\rangle[i,j]\langle 2,3,\ldots,N\rangle\Delta_1.$$

Case 1. $i>1$

Then $\langle 2,3,\ldots,N\rangle[i,j]\subset\langle 2,3,\ldots,N\rangle$ and we are done.

Case 2. $i=1,\ j=2$

Then

$$\begin{aligned}\langle 2,3,\ldots,N\rangle[1,2]&=[2,3]\cdots[2,N]\langle 3,4,\ldots,N\rangle\Delta_2\cdot[1,2]\\&\qquad\text{(by inductive hypothesis)}\\&\subset[2,3]\cdots[2,N]\langle 3,4,\ldots,N\rangle[1,2]\Delta(1,2,\ldots,N)\\&\qquad\text{by b(iii)}\\&\subset[2,3]\cdots[2,N][1,2]\langle 3,4,\ldots,N\rangle\Delta(1,2,\ldots,N)\end{aligned}$$

Now $[2,k][1,2]\subset\langle 1,2,k\rangle\subset[1,k][1,2][2,k]\Delta_2\Delta_1$ hence

$$\begin{aligned}[2,3]\cdots[2,N][1,2]&\subset[2,3]\cdots[2,N-1]\cdot[1,N][1,2][2,N]\Delta_2\Delta_1\\&\subset[1,N]\cdot[2,3]\cdots[2,N-1][1,2]\cdot\langle 2,3,\ldots,N\rangle\Delta_1\\&\subset[1,N][2,3]\cdots[2,N-2][1,N-1][1,2]\\&\qquad\times[2,N-1]\Delta_2\Delta_1\langle 2,3,\ldots,N\rangle\Delta_1\\&\subset[1,N][1,N-1]\cdot[2,3]\cdots[2,N-2]\\&\qquad\times[1,2]\cdot\langle 2,3,\ldots,N\rangle\Delta_1\\&\ \ \vdots\\&\subset[1,N][1,N-1]\cdots[1,2]\langle 2,3,\ldots,N\rangle\Delta_1\end{aligned}$$

Finally

$$\begin{aligned}[1,2]\cdots[1,N][1,N-1]\cdots[1,2]&\subset[1,2]\langle 1,3,\ldots,N\rangle[1,2]\\&\subset[1,2][1,3]\cdots[1,N]\\&\qquad\times\langle 3,4,\ldots,N\rangle\Delta(1,3,4,\ldots,N)\cdot[1,2]\\&\subset[1,2][1,3]\cdots[1,N][1,2]\langle 2,3,\ldots\ldots,N\rangle\Delta_1\\&\subset[1,2][1,3]\cdots[1,N-1]\cdot[1,2]\\&\qquad\times[1,N]\langle 2,3,\ldots,N\rangle\Delta_1\\&\subset[1,2][1,3]\cdots[1,N-2]\cdot[1,2][1,N-1]\\&\qquad\times[2,N-1]\Delta(2,N-1)\Delta_1[1,N]\\&\qquad\qquad\times\langle 2,3,\ldots,N\rangle\Delta_1\\&\subset[1,2]\cdots[1,N-2][1,2]\cdot[1,N-1][1,N]\\&\qquad\times\langle 2,3,\ldots,N\rangle\Delta_1\quad\text{by b(vii)}\\&\ \ \vdots\end{aligned}$$

Hence w lies in the group (1).

Case 3. $i=1, j>2$

Then

$$\begin{aligned}
\langle 2,3,\dots,N\rangle[1,j] &= [2,3]\cdots[2,N]\langle 3,4,\dots,N\rangle\Delta_2\cdot[1,j]\\
&\subset [2,3]\cdots[2,N]\langle 3,4,\dots,N\rangle[1,j]\cdot\Delta_1\Delta_2 \quad \text{by b(iii)}\\
&\subset [2,3]\cdots[2,N]\langle 1,3,4,\dots,N\rangle\Delta_2\Delta_1\\
&\subset [2,3]\cdots[2,N][1,3]\cdots[1,N]\langle 3,4,\dots,N\rangle\Delta(1,2,\dots,N)
\end{aligned}$$

And

$$\begin{aligned}
[2,3]\cdots[2,N][1,j] &= [2,3]\cdots[2,j-1][2,j][1,j][2,j+1]\cdots[2,N]\\
&\subset [2,3]\cdots[2,j-1]\cdot[1,j][1,2]\\
&\quad\times[2,j]\Delta_2\Delta_1\cdot[2,j+1]\cdots[2,N]\\
&\subset [1,j]\cdot[2,3]\cdots[2,j-1][1,2]\langle 2,3,\dots,N\rangle\Delta_1
\end{aligned}$$

Finally

$$\begin{aligned}
[1,2]\cdots[1,N][1,j] &\subset [1,2]\cdots[1,N-1]\cdot[1,j][1,N][j,N]\Delta(j,N)\Delta_1\\
&\subset [1,2]\cdots[1,N-2][1,j][1,N-1][j,N-1]\Delta(j,N-1)\Delta_1\\
&\quad\times[1,N][j,N]\Delta(j,N)\Delta_1\\
&\subset [1,2]\cdots[1,N-2][1,j][1,N-1][1,N]\langle 2,3,\dots,N\rangle\Delta_1\\
&\quad\vdots\\
&\subset [1,2]\cdots[1,N]\langle 2,3,\dots,N\rangle\Delta_1
\end{aligned}$$

Hence

$$w\in[1,2]\cdots[1,N]\langle 2,3,\dots,N\rangle\Delta_1\cdot[2,3]\cdots[2,j-1][1,2]\langle 2,\dots,N\rangle\Delta_1$$

and then by Case 2, w lies in the group (1). This completes the proof of Theorem 1.

3. **Presentation for** Λ_n. The presentation for Λ_n can now be derived from the presentation for Γ_n and the action of P_n on the generators:

$$A_iB_{jk}A_i=\begin{cases} B_{jk} & \text{if } i,\ j,\ k \text{ are distinct}\\ B_{jk}^{-1} & \text{if } \quad i=j \quad \text{or} \quad i=k\end{cases}$$

Thus we have Λ_n is the group with generators A_i, B_{ij}, $1\le i, j\le n$, $i\ne j$, and relations:

(a) $B_{ij}\leftrightarrow B_{ik}$
(b) $B_{ji}\leftrightarrow B_{ki}$
(c) $[B_{ij}, B_{jk}]=B_{ik}^2$
(d) $[B_{ij}B_{kj}, B_{ji}B_{jk}^{-1})=(B_{ik}B_{ki}^{-1})^2$, $[B_{ji}B_{jk}, B_{ij}B_{kj}^{-1}]=(B_{ik}B_{ki}^{-1})^2$
(e) $B_{ij}\leftrightarrow B_{kl}$

(f) $A_i^2 = 1$
(g) $A_i \leftrightarrow A_j$
(h) $A_i B_{ij} A_i = B_{ij}^{-1}$
(i) $A_i B_{ji} A_i = B_{ji}^{-1}$
(j) $A_i \leftrightarrow B_{jk}$

for distinct indices i, j, k, l. For the group Λ_n^+, as a quotient group of Λ_n, add the relation

(k) $A_1 A_2 \cdots A_n = 1$.

REFERENCE

1. H. S. M. Coxeter and W. O. J. Moser, Generators and relations for discrete groups, Ergebnisse der Mathematik und ihrer Grenzgebiete, Volume 14, Springer 1972.

AUSTRALIAN NATIONAL UNIVERSITY AND
UNIVERSITY OF BRITISH COLUMBIA

Canad. Math. Bull. Vol. **25** (2), 1982

CYCLOTOMIC SPLITTING FIELDS

BY
R. A. MOLLIN

ABSTRACT. Let D be a division algebra whose class $[D]$ is in $B(K)$, the Brauer group of an algebraic number field K. If $[D \otimes_K L]$ is the trivial class in $B(L)$, then we say that L is a *splitting field* for D or L *splits* D. The splitting fields in D of smallest dimension are the *maximal subfields* of D. Although there are infinitely many maximal subfields of D which are cyclic extensions of K; from the perspective of the *Schur Subgroup* $S(K)$ of $B(K)$ the natural splitting fields are the cyclotomic ones. In (*Cyclotomic Splitting Fields*, Proc. Amer. Math. Soc. **25** (1970), 630–633) there are errors which have led to the main result of this paper, namely to provide necessary and sufficient conditions for (D) in $S(K)$ to have a maximal subfield which is a cyclic cyclotomic extension of K, a finite abelian extension of Q. A similar result is provided for quaternion division algebras in $B(K)$.

Introduction. In this paper we are interested in cyclic cyclotomic splitting fields for division algebras. In [6, Th. 4.7, p. 757], [7, Th. 4.2, p. 207], [8, Th. 4, p. 113] and [9] we demonstrated the importance of obtaining such maximal subfields from the point of view of explicit construction of crossed product division algebras. In [11] M. Schacher gave examples of division algebras D of exponent p for each prime p with $[D] \in B(K)$ such that D does *not* have a maximal subfield which is imbedded in a cyclotomic extension of K. However there are errors in the main results of [11] which have led us to formulate the following.

In this paper we present necessary and sufficient conditions for a division algebra D with $[D] \in S(K)$ to have a maximal subfield which is a cyclic cyclotomic extension of K where K is a finite abelian extension of the field Q of rational numbers.

Moreover, for $[D] \in B(K)$ with K a finite non-real abelian extension of Q where D is a quaternion division algebra we provide necessary and sufficient conditions for D to have a maximal subfield which is a cyclic cyclotomic extension of K.

1. Notation and preliminaries. Let K be a field of characteristic zero. The Schur group $S(K)$ may be described as consisting of those equivalence classes

---

Received by the editors September 16, 1980.
AMS (MOS) Subject Classification Numbers: Primary: 16A26; Secondary: 16A65.
Key Words and Phrases: Cyclotomic field, splitting field, division algebra, maximal subfield.
The author's research is supported by an N.S.E.R.C. University Research Fellowship.

in $B(K)$ which contain a simple component of the group algebra KG for some finite group G. For basic results concerning $S(K)$ the reader is referred to [14].

When K is an algebraic number field the elements $[A] \in B(K)$ are uniquely characterized by their Hasse invariants. A certain subgroup of $B(K)$ has a particularly nice relationship between these invariants. We describe it as follows:

Let K be a finite abelian extension of Q. $U(K)$, called the *absolute uniform distribution group for* K, denotes the subgroup of $B(K)$ consisting of those equivalence classes $[A]$ such that:

(i) If the index of A is n then ε_n is in K, where ε_n denotes a primitive nth root of unity, and

(ii) If P is a K-prime above the rational prime p and $\sigma \in G(K/Q)$, the Galois group of K over Q, with $\varepsilon_n^\sigma = \varepsilon_n^b$ then the *Hasse P-invariant* of A satisfies:

$$\mathrm{inv}_P A \equiv b \, \mathrm{inv}_{P^\sigma} A \pmod 1.$$

If $[A] \in U(K)$ and P and P' are K-primes above the rational prime p then $A \otimes_K K_P$ and $A \otimes_K K_{P'}$ have the same index, where K_P denotes the completion of K at p. The common values of the indicies $A \otimes_K K_P$ for all K-primes P above p is called the *p-local index* of A, denoted $\mathrm{ind}_p A$.

We studied the relationship between $S(K)$ and $U(K)$ of which it is a subgroup in [4]–[9].

If $[A] \in B(K)$ and $\mathrm{inv}_P A > 0$ for a$K-$prime P then we say that P is *ramified in* A, (see [10, p. 272]). Since we shall be concerned with K/Q finite abelian then we may say that p is ramified in A where p is the rational prime below P, whenever $\mathrm{inv}_P A > 0$ for some K-prime P.

The *norm-residue symbol* at P is denoted $(*, *)_P$ and the *Legendre symbol* is denoted $(/)$.

Throughout the remainder of the paper we shall be concerned with finite abelian extensions K of Q. A field extension K of F shall be denoted K/F. Since the decomposition of an F-prime in K essentially depends on the rational prime q which sits below it then we shall write F_q to denote the completion of K at an F-prime above q. Similarly K_q shall denote the completion of a K-prime above the given F-prime.

If G is a group and p is a prime then G_p shall denote the Sylow p-subgroup of G. If $m = p^a t$ where p and t are relatively prime then $|m|_p = p^a$, i.e. $|m|_p$ denotes the highest power of p dividing the integer m.

A crossed product algebra is denoted by $(L/K, \beta)$. This is the central simple K-algebra having L-basis u_τ with $\tau \in G(L/K) = G$ subject to:

$$u_\tau u_\sigma = \beta(\tau, \alpha) u_{\tau\sigma}, \qquad \tau, \sigma \in G$$

and

$$u_\tau x = x^\tau u_\tau \quad \text{for} \quad x \in L^*.$$

Moreover a crossed product of the form $(K(\varepsilon)/K, \beta)$ where ε is a root of unity and the values of β are roots of unity in $K(\varepsilon)$ are called *cyclotomic algebras*. These are the algebras which characterize $S(K)$, (see [14]).

When G is cyclic then $(L/K, \beta)$ denotes the cross product in which:

$$u_\tau^i = \begin{cases} u_{\tau^i} & \text{if} \quad 1 \le i < |L:K| \\ \beta & \text{if} \quad i = |L:K|. \end{cases}$$

For further information on crossed products the reader is referred to [10]. Finally equivalence in $B(K)$ will be denoted by $\sim$.

2. Splitting fields for quaternion division algebras.

Let K/Q be finite abelian and let D be a division algebra with $[D] \in B(K)$. We note that to ask whether D has a maximal subfield which can be imbedded in a cyclotomic extension of K is rendered, by the Kronecker-Weber theorem, to be equivalent to asking whether D has a maximal subfield which is abelian over Q. We commence by asking whether a quaternion division algebra D has a maximal subfield abelian over Q. The answer is negative in general as the following counterexample illustrates.

Let $K = Q(\sqrt{-1}, \sqrt{3})$ and let $[D] \in B(K)$ (in fact $[D] \in U(K)$), with $\operatorname{ind}_2 D = 2 = \operatorname{ind}_3 D$, and $\operatorname{ind}_p D = 1$ for all primes $p \neq 2, 3$. If a maximal subfield L of D exists such that L/Q is abelian then either:

(1) $G(L/Q) = Z_2 \oplus Z_2 \oplus Z_2$ or,

(2) $G(L/Q) = Z_2 \oplus Z_4$.

If (1) then $G(L_3/Q_3) = Z_2 \oplus Z_2 \oplus Z_2$. However, by [13, 6–5–4] this is not possible since Q_3 has only three quadratic extensions. Thus (2) holds and so one of $Q(\sqrt{-1})$, $Q(\sqrt{3})$ or $Q(\sqrt{-3})$ is imbedded in a cyclic extension of degree 4. By [1, Th. 6, p. 106], -1 must be a norm from one of these three fields. However, -1 cannot be a norm from an imaginary quadratic field. Therefore -1 must be a norm from $Q(\sqrt{3})$. Thus, by the Hasse norm theorem -1 must be a norm everywhere locally. However, $(3, -1)_3 = (-\frac{1}{3}) = -1$; i.e. -1 is not a norm from $Q_3(\sqrt{3})$, a contradiction which establishes the counterexample.

The above example is similar to [11, p. 632]. However the example therein is incorrect. We shall come back to this once we have the first result at our disposal. The following theorem provides necessary and sufficient conditions for a quaternion division algebra to have a maximal subfield abelian over Q. In what follows we shall use the term *maximal cyclic p-extension* of F in K to mean a proper subfield M of K such that $G(M/F)_p$ is cyclic, and if $F \subseteq M \subseteq N \subseteq K$ with $G(N/F)_p$ cyclic then $|N:M|_p = 1$.

THEOREM 2.1. *Let K/Q be finite non-real abelian, and let D be a quaternion division algebra with $[D] \in B(K)$. D has a maximal subfield which is abelian* *over Q if and only if for each odd prime q which ramifies in D with $G(K_q/Q_q)_2$ non-cyclic, there exists a maximal cyclic 2-extension F of Q in K such that:*

(a) -1 *is a norm in F/Q and,*

(b) *q is not completely split in F/Q.*

Proof. First we prove the necessity of (a) and (b). Suppose there exists a maximal subfield L of D such that L/Q is abelian. If q is an odd prime which ramifies in D with $G(K_q/Q_q)$ non-cyclic then by [13, 6-5-4], $G(L_q/Q_q)_2$ must be of the form $Z_{2^{m(1)}} \oplus Z_{2^{m(2)}}$ where $m(1)$ and $m(2)$ are positive integers. Thus there exist maximal cyclic 2-extension M_i of Q in L with $|M_i : Q|_2 \geq 2^{m(i)}$ for $i = 1, 2$. One of M_1 or M_2 is not contained in K, say $M = M_1$. Therefore, if $M \cap K = F$ then $|M:F| = 2$ and F is a maximal cyclic 2-extension of Q in K. By [1, Th. 6, p. 106] -1 is a norm in F/Q. Moreover since $G(K_q/Q_q)_2$ is non-cyclic then q is not completely split in F.

Conversely suppose (a) and (b) hold. Let $q(i)$ for $i = 1, 2, \ldots, m$ be all rational primes which ramify in D but do not ramify in K/Q. Set $\alpha(i) = q(i)$ for $i = 1, 2, \ldots, m$. Since $q(i)$ is ramified in $K(\sqrt{\alpha(1)\alpha(2) \cdots \alpha(m)})/K$ then $K(\sqrt{\alpha(1)\alpha(2) \cdots \alpha(m)})$ splits D at each $q(i)$ for $i = 1, 2, \ldots, m$.

Now consider $T = \{q(m+1), q(m+2), \ldots, q(n)\}$ where $q(i)$ is odd, ramified in D and, $G(K_{q(i)}/Q_{q(i)})_2$ is not cyclic for $i = m+1, m+2, \ldots, n$. We note that $q(i)$ splits in $K(\sqrt{\alpha(1)\alpha(2) \cdots \alpha(m)})/K$ for $i = m+1, m+2, \ldots, n$ since otherwise we would have a degree 8 extension of Q_q with Galois group of the form $Z_2 \oplus Z_2 \oplus Z_2$ which would contradict [13, 6-5-4]. Now, by hypothesis, for each $q(i) \in T$ there exists a maximal cyclic 2-extension $F^{(i)}$ of Q in K satisfying (a) and (b). Not all such $F^{(i)}$ are necessarily distinct, so we let $F^{(j)}$ for $j = m+1, \ldots, r$ with $m+1 \leq r \leq n$ be all distinct such fields. Now we rearrange the elements of T as follows. Let

$$R(j) = \{q(i, j) \in T : i = m(j-1)+1, \ldots, m(j) \text{ with } m(m) = m \text{ and } m(r) = n\}$$

where $j = m+1, \ldots, r$, be the set of all elements of T which are not completely split in $F^{(j)}$ and which do not already appear in $R(k)$ for $m+1 \leq k < j$. Since $G(K_{q(i)}/Q_{q(i)})_2$ is not cyclic for $i = m+1, \ldots, n$ then it is possible to ensure as well that $q(i, j)$ is completely split in $F^{(h)}$ for all $h \neq j$. Now, by hypothesis -1 is a norm from $F^{(j)}$ for $j = m+1, \ldots, r$. By [1, Th. 6, p. 106] $F^{(j)}$ is contained in $M^{(j)}$ where $|M^{(j)} : F^{(j)}| = 2$ and $M^{(j)}$ is cyclic over Q. Since $F^{(j)}$ is a maximal cyclic 2-extension Q in K then $|M^{(j)}K : K| = 2$ and by Kummer theory $M^{(j)}K = K\sqrt{\beta_j})$ for some $\beta_j \in K^*$. We note that since K/Q is abelian and $M^{(j)}/Q$ is cyclic then $M^{(j)}K/Q$ is abelian. Therefore by Kronecker-Weber $K(\sqrt{\beta_j})$ is contained in a cyclotomic extension of K. Now we choose $\alpha(m(j-1)+1 = \beta_j$ and $\alpha(m(j-1)+2) = \cdots = \alpha(m(j)) = 1$ for $j = m+1, \ldots, r$, and set $\alpha(r+1) = \alpha(r+2) = \cdots = \alpha(n) = 1$.

Finally we consider those remaining $q(i)$ for $i=n+1,\ldots,s$ which ramify in D. First we consider those $q(i)$ which are either odd or for which $G(K_{q(i)}/Q_{q(i)})_2$ is cyclic. If $q(i)$ does not split in $K(\sqrt{\alpha(1)\cdots\alpha(i-1)})/K$ then set $\alpha(i)=1$. Otherwise choose a prime $p(i)$ which is relatively prime to the discriminant of $K(\sqrt{\alpha(1)\cdots\alpha(i-1)})$ and such that $q(i)$ is inert in $Q(\sqrt{p(i)})$ while $q(j)$ splits in $Q(\sqrt{p(i)})$ for all $j<i$. Such $p(i)$ exist by Chinese remainder theorem considerations.

The only possible remaining case is $q(s)=2$ where $G(K_2/Q_2)_2$ is not cyclic. If 2 does not split in $K(\sqrt{\alpha(1)\cdots\alpha(s-1)})/K$ then set $\alpha(s)=1$. Let $\gamma=\alpha(1)\cdots\alpha(s-1)$.

Otherwise if $\sqrt{-1}\notin K$ set $\alpha(s)=\sqrt{-1}$ or $\alpha(s)=\sqrt{2}$ according as 2 is nonsplit in $K(\sqrt{-1\gamma})/K$ or $K(\sqrt{2\gamma})/K$. We note that 2 cannot be split in $K(\gamma)/K$, $K(\sqrt{-1\gamma})/K$ and $K(\sqrt{2\gamma})/K$ since in that case 2 could be split in $K(\varepsilon_8)/K$ contradicting $\sqrt{-1}\notin K$. If $\sqrt{-1}\in K$ and ε_{2^a} for $a>1$ is the largest 2-power root of unity in K then 2 does not split in $K(\varepsilon_{2^{a+1}}\gamma)/K$. In this case set $\alpha(s)=\varepsilon_{2^a}$.

By construction $L=K(\sqrt{\alpha(1)\cdots\alpha(s)})$ splits D at all primes which ramify in D, and L is abelian over Q. It follows that L is a maximal subfield of D which secures the theorem. Q.E.D.

We isolate a special case of Theorem 2.1 since it has a bearing on [11].

COROLLARY 2.2. *Let K be a biquadratic extension of Q. Then every quaternion division algebra in $U(K)$ has a maximal subfield which can be imbedded in a cyclotomic extension of Q if and only if either:*

(a) $|K_q:Q_q|=4$ *for at most one prime q, or*

(b) -1 *is a norm from one of the quadratic subfields of K.*

For example, for $K=Q(\sqrt{-1},\sqrt{7})$ then only prime q with $|K_q:Q_q|=4$ is $q=7$. Therefore by Corollary 2.2 every quaternion division algebra in $U(K)$ has a maximal subfield which is abelian over Q. This shows that the example [11, p. 632] is false, and that no such algebra $[D]\in U(K)$ can be found. The error stems from Schacher's claim that "... one easily checks that $G(K_2/Q_2)=G(K_7/Q_7)=Z_2\oplus Z_2$." In fact one checks that $G(K_2/Q_2)=Z_2$ since 2 splits in $Q(\sqrt{-7})$.

That K is restricted to being non-real in Theorem 2.1 is a result of problems which occur at 2 and the infinite rational primes. Similar problems were encountered in [8, Th. 1, p. 108] but resolved by a suitable restriction [8, Th. 2, p. 112]. In §3 we shall overcome the problem by considering a special subgroup $S(K)$ of $B(K)$.

3. Splitting fields and $S(K)$. In this section we restrict our attention to division algebras D with $[D]\in S(K)$ where K/Q is finite abelian.

In [8] we considered the following situation. Let χ be a complex irreducible character of a finite group G of exponent n. Let $A(\chi, Q)$ denote the simple component of QG corresponding to χ. We note that $[A(\chi, Q)] \in S(Q(\chi))$. R. Brauer's well known theorem which states that $Q(\varepsilon_n)$ is a splitting field for χ, inspired the following demanding question: Does the division algebra underlying $A(\chi, Q)$ have a maximal subfield L contained in $Q(\varepsilon_n)$? In general the answer is negative, and in [8] we provided sufficient conditions for such an L to exist. However, for each result which we obtained we were able to find counterexamples to the necessity of such conditions. In this paper we relax the demands on L. We merely require that L/Q be abelian, i.e. L may be imbedded in *any* cyclotomic extension of Q. In [2] B. Fein found counterexamples to the existence of such an L for each prime p. We now present for the first time necessary and sufficient conditions for such an L to exist.

THEOREM 3.1. *Let K/Q be finite abelian and let D be a division algebra of index m with $[D] \in S(K)$. D has a maximal subfield cyclic over K and abelian over Q if and only if for each odd prime q which ramifies in D and for each prime p dividing m with $G(K_q/Q_q)_p$ non-cyclic there exists a maximal cyclic p-extension F of $Q(\varepsilon_{p^c})$ in K where $|\mathrm{ind}_q D|_p = p^c$ such that*

(a) ε_{p^c} is a norm in $F/Q(\varepsilon_{p^c})$ and

(b) q is not completely split in $F/Q(\varepsilon_{p^c})$.

Proof. We note that if $m = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ where the p_i's are distinct primes then $D \sim D_1 \otimes \cdots \otimes D_r$ in $S(K)$ where the index of D_i is $p_i^{a_i}$ for $i = 1, 2, \ldots, r$. Thus it follows that we may assume without loss of generality that $m = p^b$.

First we prove the necessity of (a) and (b). Suppose $G(K_q/Q_q)_p$ is non-cyclic for odd q with $\mathrm{ind}_q D = p^c$ where $c \leq b$. By [5, Th. 1.1, p. 273] $q \equiv 1 \pmod{p^c}$. If D has a maximal subfield abelian over Q and cyclic over K then by [13, 6-5-4] $G(L_q/Q_q)_p$ is forced to be of the form $Z_{p^{m(1)}} \oplus Z_{p^{m(2)}}$ where one of $m(1)$ or $m(2)$ is greater than c, say $m = m(1) > c$ and $m(2) > 0$. Therefore there exists a maximal cyclic p-extension M of $Q(\varepsilon_{p^c})$ in L with $|M : Q|_p > p^c$. Thus $M \cap K = F$ is a maximal cyclic p-extension of $Q(\varepsilon_{p^c})$ in K with $|M : F|_p = p^c$. By [1, Th. 6, p. 106] ε_{p^c} is a norm in $F/Q(\varepsilon_{p^c})$ and since $m > c$ then q is not completely split in $F/Q(\varepsilon_{p^c})$. This establishes the necessity.

Suppose $\mathrm{ind}_{q(i)} D = p^{c(i)}$ with $q(i)$ unramified in K/Q for $i = 1, 2, \ldots, m$. That there exists a subfield L_i of $K(\varepsilon_{q(i)})$ with $|L : K| = p^{c(i)}$ can be verified by exactly the same argument as in [8, Th. 1, p. 109]. By [14, Prop. 6.2, p. 89] we have $\varepsilon_{p^{c(i)}}$ is in K and so $L_i = K(\gamma(i))$ where $\gamma(i)^{p^{c(i)}} \in K$ for $i = 1, 2, \ldots, m$. Since $q(i)$ ramifies in L_i/K then L_i splits D at $q(i)$ for $i = 1, 2, \ldots, m$.

Now consider those primes $q(i)$ with $\mathrm{ind}_{q(i)} D = p^{c(i)}$ for $i = m+1, \ldots, n$ such that $q(i)$ is odd, and $G(K_{q(i)}/Q_{q(i)})_p$ is non-cyclic. Using (a) and (b) of the hypothesis we can use exactly the same kind of argument as in Theorem 2.1 to obtain fields L_i abelian over Q and cyclic over K such that L_i splits D at $q(i)$ for $i = m+1, \ldots, n$. Set $L_i = K(\gamma(i))$.

Now we consider the remaining odd primes $q(i)$ for $i=n+1,\ldots,s$ with $\mathrm{ind}_{q(i)} D=p^{c(i)}$. If $q(i)$ does not split in $K(\gamma(1)\cdots\gamma(i-1))$ then set $\gamma(i)=1$. Otherwise by [3, Prop. 5.2, p. 275] we may choose a prime $p_i\equiv 1 \pmod p$ such that $q(i)$ is not a pth power modulo p_i but $q(1), q(2),\ldots, q(i-1)$ are $p^{c(i)}$-th powers modulo p_i. Thus there exists a field M_i contained in $K(\varepsilon_{p_i})$ such that $|M_i:K|=p^{c(i)}$ with $q(i)$ inert in M_i/K and $q(j)$ completely split in M_i/K for all $j<i$. By Kummer theory $M_i=K(\gamma_{(i)})$ where $\gamma_{(i)}^{p^{c(i)}}\in K^*$. Set $\gamma'=\gamma(1)\cdots\gamma(s)$. Hence $K(\gamma(1)\cdots\gamma(s))$ splits D at $q(i)$ for $i=1,\ldots,s$.

If $\mathrm{ind}_2 D=2$ and $\mathrm{ind}_\infty D=1$ then set $\gamma'=\alpha$ if 2 does not split in $K(\gamma')/K$ and set $\sqrt{-1}\gamma'=\alpha$ otherwise. We note that by [14, Th. 5.11(II), p. 81] 2 is ramified in $K(\sqrt{-1})/K$. Hence if 2 splits in $K(\gamma')/K$ then 2 ramifies in $K(\sqrt{-1}\gamma)/K$.

If $\mathrm{ind}_\infty D=2$ and $\mathrm{ind}_2 D=1$ then set $\gamma'=\alpha$ if $K(\gamma')$ is non-real, and set $\sqrt{-1}\gamma'=\alpha$ otherwise. Clearly $K(\alpha)$ splits D at ∞.

Suppose $\mathrm{ind}_\infty D=2=\mathrm{ind}_2 D$. If 2 is not split in $K(\gamma')/K$ and $K(\gamma')$ is not real then set $\gamma'=\alpha$. If 2 splits in $K(\gamma')/K$ then 2 ramifies in $K(\sqrt{-1}\gamma')/K$, (ibid.). In this case set $\sqrt{-1}\gamma'=\alpha$. We note that by the choice of γ' it is not possible to have the case where $K(\gamma')$ is non-real but $K(\sqrt{-1}\gamma')$ is real. Hence $K(\alpha)$ splits D at 2, and ∞.

We are left with the case where $\mathrm{ind}_2 D=\mathrm{ind}_\infty D=2$ and 2 does not split in $K(\gamma')/K$, where $K(\gamma')$ is real. Then we consider 2 cases:

(a) 2 does not split in $K(\sqrt{-2\gamma'})/K$. In this case set $\sqrt{-2\gamma'}=\alpha$.

(b) 2 splits in $K(\sqrt{-2\gamma'})/K$. Therefore 2 splits in $K(\sqrt{2})K$. Since K is real then K contains a quadratic subfield $Q(\sqrt{d})$ where d is an even square-free integer. Suppose $Q(\varepsilon_r)$ is the smallest root of unity field containing K, with $|r|_2=2^t$; $t>2$. In this case choose $\sqrt{-1}(\varepsilon_{2^{t+1}}+\varepsilon_{2^{t+1}}^{-1})\gamma'=\alpha$. By [14, Prop. 7.5, p. 103] $K(\alpha)$ is not real and 2 ramifies in $K(\alpha)/K$. Thus $K(\alpha)$ splits D at 2 and ∞.

Since $m=p^b$ then $\mathrm{ind}_{q(i)} D=p^b$ for some i. Thus $|L_i:K|=p^b$ for some i which implies $|K(\alpha):K|=p^b$. By construction $L=K(\alpha)$ splits D at each $q(i)$ for $i=0,1,\ldots,s$, L/K is cyclic, and L/Q is abelian. It follows that L is the required maximal subfield of D. Q.E.D.

Now that we have necessary and sufficient conditions for the existence of a maximal subfield L of D to be abelian over Q and cyclic over K we ask: Once we have L, is it possible to find a suitable factor set α such that $D\sim(L/K,\alpha)$ in $S(K)$? The answer is yes in general, see [10]. If, however, we require the more demanding restriction that α be a root of unity in K then the answer is negative in general. Although Yamada [14, p. 33] has shown that every element A with $[A]\in S(K)$ is equivalent to a cyclotomic algebra it is not necessarily the case that the division algebra underlying A is also cyclotomic. This is in fact what we are requiring by our more demanding restriction on α. In Mollin [9] we have provided necessary and sufficient conditions for a division algebra to be cyclotomic.

It is natural to ask whether Theorem 3.1 holds for a larger class of elements than those in $S(K)$. M. Schacher [11, Th. 1, p. 630] provides a counter-example of exponent p, one for every prime p. However, there is an error in his proof. The following is a counter-example to [11, Th. 1, p. 630].

Let q be an odd prime such that 2 is a primitive root modulo q, and let p be an odd prime such that $q \equiv 1 \pmod{p^2}$. Let K be the unique subfield of $Q(\varepsilon_q)$ which has degree p over Q. We define $[D] \in U(K)$ as follows:

$$\operatorname{ind}_2 D = 1/p \quad \text{and} \quad \operatorname{ind}_q D = 1/p \quad \text{and} \quad \operatorname{ind}_r D = 1 \quad \text{for all} \quad r \neq 2, q.$$

Since $q \equiv 1 \pmod{p^2}$ then K is contained in a subfield L of $Q(\varepsilon_q)$ such that $|L:K| = p$. Since 2 is a primitive root modulo q then $|L_2:K_2| = p$ and clearly $|L_q:K_q| = p$. Thus L is a maximal subfield of D, cyclic over K and abelian over Q, contradicting [11, Th. 1, p. 630].

The error in Schacher's proof arises essentially from one of his references, viz. Serre's [12, Prop. 5, p. 92] in which there is a misprint. Serre's result should read "$\ldots N_0(\xi) = \xi^l \ldots$" which translates in Schacher's notation to: $N_0(\xi) = \xi^p$. We see therefore, that if $q \not\equiv 1 \pmod{p^2}$ then his proof fails. We note however that if $q \not\equiv 1 \pmod{p^2}$ then, with the correct interpretation of [12, Prop. 5, p. 92] his proof would hold. Dr. Serre has informed me in a recent letter that the aforementioned misprint has been corrected in the English edition.

REFERENCES

1. E. Artin and J. Tate, *Class Field Theory* Benjamin, New York, (1968)
2. B. Fein, *Minimal Splitting Fields for Group Representations II*, Pacific J. Math, **77** (2), (1978), 445–449.
3. J. Janusz, *The Schur Group of an Algebraic Number Field*, Annals of Math., **103**, (1976), 253–281.
4. R. Mollin, *Uniform Distribution and the Schur Subgroup*, J. Algebra **42**, (1976), 261–277.
5. R. Mollin, *Algebras with Uniformly Distributed Invariants*, J. Algebra, **44**, (1977), 271–282.
6. R. Mollin, *U(K) for a Quadratic Field K*, Communications in Algebra, **4**(8), (1976), 747–759.
7. R. Mollin, *Uniform Distribution Classified*, Math. Zeitschrift, **165**, (1979), 199–21.
8. R. Mollin, *Splitting Fields and Group Characters J. reine angew Math.* **315**, (1980), 107–114.
9. R. Mollin, *Cyclotomic Division Algebras*. To appear in Canadian J. Math.
10. I. Reiner, *Maximal Orders*, Academic Press, N.Y. (1975).
11. M. Schacher, *Cyclotomic Splitting Fields*, Proc. Amer. Math. Soc., **25**, (1970), 630–633.
12. J. P. Serre, Corps Locaux, Actualités Sci. Indust., No. 1296, Hermann, Paris, 1962.
13. E. Weiss, *Algebraic Number Theory*, McGraw-Hill, New York, (1963).
14. T. Yamada, *The Schur Subgroup of the Brauer Group*, Lecture Notes in Math., No. 397, Springer-Verlag, (1974).

DEPARTMENT OF MATHEMATICS AND STATISTICS
QUEEN'S UNIVERSITY
KINGSTON, ONTARIO
K7L 3N6.

Canad. Math. Bull. Vol. **25** (2), 1982

SYSTEMS OF BRIOT-BOUQUET EQUATIONS WITH ANALYTIC SOLUTIONS

BY

DAVID WESTREICH AND ESTER PODOLAK

ABSTRACT. In this note we use functional analysis arguments to prove the existence of families of analytic solutions for the singular system of complex ordinary differential equations $zW' = h(z, W)$.

Introduction. We consider the singular system of complex ordinary differential equations $zW' = h(z, W)$ where h is a holomorphic complex-valued vector function of (z, W), the z-space has one complex dimension and the W-space is complex n-dimensional. We prove the existence of families of analytic solutions indexed by the independent eigenvectors of the matrix $h_W(0, 0)$ corresponding to the integer eigenvalue m.

Such systems were first studied by C. C. A. Briot and J. C. Bouquet [1] for the one-dimensional case. One can show that in this case the equation has a one-parameter family of analytic solutions [6, p. 116]. Further results appear in [4] where the approach taken has been to obtain formal power series solutions and then prove convergence. An alternative approach was taken in [3] using a contraction mapping to obtain the existence of families of analytic solutions in the n-dimensional case depending on solutions of an arbitrary polynomial equation. We obtain similar, though more explicit results using a direct functional analysis method.

Main Results. Consider the system of Briot-Bouquet equations

$$zW' = h(z, W), \qquad h(0,0) = 0 \qquad \left(' = \frac{d}{dz}\right) \tag{1}$$

where z is a complex variable, W an n-vector and $h(z, W)$ is a map of $\mathbb{C}^n \to \mathbb{C}^n$ whose components are holomorphic functions of (z, W). We wish to find a nontrivial solution of Eq. (1) such that each of its components is an analytic function.

To this end we let $h_W(z, W)$ denote the Jacobian matrix of $h(z, W)$ with respect to W. Suppose that $h_W(0, 0)$ has a positive integer eigenvalue and let m

---

Received by the editors July 4, 1979 and in revised form, March 24, 1980, July 24, 1980, August 28, 1980 and January 20, 1981.

AMS subject classification Number 34A20.

be the largest positive integer eigenvalue of $h_W(0, 0)$. We shall also assume that

$$h(z, 0) = o(|z|^m).$$

With the above assumptions we can show

THEOREM A. *There exists a neighborhood N of 0 in $\mathbb{C}^n$ such that for each eigenvector $C \in N$ of $h_W(0, 0)$ corresponding to m there is an analytic solution*

$$W(z) = Cz^m + \sum_{i=m+1}^{\infty} C_i(C)z^i, \qquad 0 \le |z| \le 1$$

of Eq. (1).

Proof. Let $A = h_W(0, 0)$ and $H(z, W) = h(z, W) - AW$. Then let B_m be the space of functions which are analytic in the circle $|z| < 1$, continuous on $|z| \le 1$ and have power series expansions of the form $\sum_{i=m+1}^{\infty} C_i z^i$. We will define $B_m^{(1)}$ to be the subspace of B_m consisting of those functions whose derivatives are also continuous on $|z| \le 1$. Then set

$$\mathscr{B}_m = \{(f_1(z), \ldots, f_n(z)) \mid f_i \in B_m, i = 1, \ldots, n\}$$

and

$$\mathscr{B}_m^{(1)} = \{(f_1(z), \ldots, f_n(z)) \mid f_i \in B_m^{(1)}, i = 1, \ldots, n\}$$

with the norms

$$\|f\|_{\mathscr{B}_m} = \sup_{\substack{|z| \le 1 \\ 1 \le i \le n}} |f_i(z)| \quad \text{and} \quad \|f\|_{\mathscr{B}_m^{(1)}} = \sup_{\substack{|z| \le 1 \\ 1 \le i \le n}} |f_i'(z)|.$$

It is readily verified that the operator $T\colon \mathscr{B}_m^{(1)} \to \mathscr{B}_m$ defined by

$$TW = zW'(z) - AW(z)$$

is a 1-1 map. Moreover

$$H(z, Cz^m + W(z)) \colon C^n \times \mathscr{B}_m^{(1)} \to \mathscr{B}_m.$$

Thus to solve Eq. (1) it suffices to solve the equation

$$zW'(z) = AW(z) + H(z, Cz^m + W(z))$$

for C an eigenvector of A corresponding to m and $W \in \mathscr{B}_m^{(1)}$.

As $f(z) = \int_0^z f'(w)\,dw$ for $f \in B_m^{(1)}$ it follows from the Arzela-Ascoli theorem that the injection map of $\mathscr{B}_m^{(1)} \to \mathscr{B}_m$ is compact. Moreover it is clear that the operator defined by $zW'(z)$ is a bounded 1-1 map of $\mathscr{B}_m^{(1)}$ onto $\mathscr{B}_m$ and is thus invertible. Hence by the stability theorem for Fredholm operators [5, p. 238] it follows that T is a linear homeomorphism of $\mathscr{B}_m^{(1)}$ onto $\mathscr{B}_m$. Therefore by the implicit function theorem [2, p. 265] for each sufficiently small eigenvector C there exists a unique $W_C(z)$ such that $zW_C'(z) = AW_C(z) + H(z, Cz^m + W_C(z))$. Thus our theorem follows.

When the nonlinear part of $h(z, W)$ is of sufficiently high order one need not assume m is a maximal integer eigenvalue to assure the existence of analytic solutions. Indeed, suppose now that m is a positive integer eigenvalue of $h_W(0,0)$ and k is the largest positive integer such that km is an eigenvalue of $h_W(0,0)$. Then we can readily show

THEOREM B. *Let m and k be as above and suppose that*

$$h(z, W) - h_W(0,0)W = o(|z|^{km} + |W|^{km}).$$

Then there exists a neighborhood N of 0 in $\mathbb{C}^n$ such that for each eigenvector $C \in N$ of $h_W(0,0)$ corresponding to m there is an analytic solution

$$W(z) = Cz^m + \sum_{i=km+1}^{\infty} C_i(C)z^i \qquad 0 \le |z| \le 1$$

of Eq. (1).

Proof. It is clear that

$$h\left(z, Cz^m + \sum_{i=km+1}^{\infty} C_i z^i\right) - ACz^m = \sum_{i=km+1}^{\infty} B_i z^i.$$

Hence if we replace B_m by B_{km} and repeat the arguments used in proving Theorem *A* our Theorem follows.

If $h(z, W)$ does not depend on z, so that Eq. (1) has the form

$$zW' = h(W), \qquad h(0) = 0 \tag{2}$$

then a stronger result obtains. Suppose that m is a positive integer eigenvalue of $h_W(0)$ and no other positive integer multiple of m is an eigenvalue of $h_W(0)$. Then

THEOREM C. *There exists a neighbourhood N of 0 in C^n such that for each eigenvector $C \in N$ of A corresponding to m there is an analytic solution*

$$W(z) = Cz^m + \sum_{i=2}^{\infty} C_{mi}(C)z^{mi} \qquad 0 \le |z| \le 1.$$

of Eq. (2).

Proof. The proof is almost the same as that of Theorem A. Just take B_m to be the space of analytic functions with power series of the form $\sum_{i=2}^{\infty} C_{mi}z^{mi}$.

ACKNOWLEDGEMENT. The authors wish to thank the referee for his helpful comments.

REFERENCES

1. C. C. A. Briot and J. C. Bouquet, *Recherches sur les proprietes des fonctions definies par des equations differentielles*, J. Ecole Poly. Tome **21** (1856), 133–197.

2. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

3. W. A. Harris, Jr., *Holomorphic solutions of nonlinear differential equations of singular points*, SIAM Studies in Appl. Math. **5**, (1969), 184–187.

4. M. Iwano, *Analytic expressions for bounded solutions of nonlinear ordinary differential equations with Briot-Bouquet type singularity*, Funkcialag Ekracioj, **12** (1969), 41–88.

5. T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag Berlin (1966).

6. G. Sansone and R. Conti, *Nonlinear Differential Equations*, MacMillan, New York (1964).

DEPARTMENT OF MATHEMATICS,
BEN GURION UNIVERSITY OF THE NEGEV,
BEER-SHEVA, ISRAEL.

DEPARTMENT OF MATHEMATICS,
BEER-SHEVA, ISRAEL.
RAMAT-GAN, ISRAEL.

Canad. Math. Bull. Vol. **25** (2), 1982

A NEW PROOF OF THE SNAKE THEOREM

BY
LEE L. KEENER

ABSTRACT. The Snake Theorem (terminology of Krein), due to Karlin in its original form, has been periodically improved. The theorem shows under appropriate conditions the existence of a function p^* from a Tchebycheff space T, with a graph that alternately "touches" the graphs of functions f and g where $f<g$ and $f \leq p^* \leq g$ on a compact interval $[a, b]$. The number of "touchings" depends upon the dimension of T. In this paper the conditions assumed are not the weakest known (see Gopinath and Kurshan, J. of Approximation Theory **21** (1977), 151–173), but the apparently new proof offered is elementary and fairly short. f and g are not assumed continuous.

1. One of the most beautiful theorems of analysis is the oscillation theorem due to Karlin [3], [4], descriptively termed the "Snake Theorem" by Krein and Nudel'man [6]. A number of alternate proofs of this theorem have appeared in the literature, notably those of Krein and Nudel'man [6], Pinkus [7], and Gopinath and Kurshan [2]. In fact the last authors prove a generalized oscillation theorem wherein Karlin's original hypotheses are substantially weakened. The snake theorem is usually considered a "deep" theorem and the proofs in the literature are either long and complex [2], or rely upon other "deep" theorems such as the Brouwer Fixed Point Theorem [4]. In this note we use a different technique to prove the theorem in an elementary way. Our hypotheses are weaker than those of [4] or [7], but stronger than those of [2]. The reader is referred to [4] for the definitions of Tchebycheff space, nodal zero and related concepts. We assume the uniform topology is used on $C[a, b]$.

2. DEFINITION. Two functions u and v, defined on $[a, b]$, are said to touch at x_0 in $[a, b]$ if there is a sequence $\langle x_i \rangle$ in $[a, b]$ such that $x_i \rightarrow x_0$ and $u(x_i)-v(x_i) \rightarrow 0$.

THEOREM 1. *Let f and g be two functions defined on $[a, b]$ and let T be an n-dimensional Tchebycheff space of continuous functions on $[a, b]$. Assume there is a function w in T and $\varepsilon>0$ such that $f(x)+\varepsilon \leq w(x) \leq g(x)-\varepsilon$ for all $x \varepsilon [a, b]$. Then there is an element p^* in T and points $x_1<x_2<\cdots<x_n$ in $[a, b]$*

Received by the editors March 25, 1980.

AMS (1980) Classification: 41A05; 41A99.

Key words: Tchebycheff Space, Snake Theorem.

such that

(a) $f(x) \le p^*(x) \le g(x)$ *for all* x in $[a, b]$

(b) f *touches* p^* at x_i *for* i *odd*, $1 \le i \le n$.

(c) g *touches* p^* at x_i *for* i *even*, $2 \le i \le n$.

Furthermore, there is a function q^* *in* T *that satisfies conditions* (a′), (b′), (c′) *obtained from* (a), (b), (c) *by replacing* p^* *by* q^* *and interchanging* f *and* g *in* (b) *and* (c). *The functions* p^* *and* q^* *are the only functions in* T *satisfying* (a), (b), (c) *and* (a′), (b′), (c′) respectively.

Proof. We show only the existence of p^*, the q^* case being similar. The uniqueness of p^* and q^* can be shown by standard zero-counting arguments; (see [4], p. 70). Without loss of generality we may assume $f < 0$, $g > 0$ with both f and g bounded away from zero. Let $M \subset T$ be the set of all functions p in T for which there exist n points in $[a, b]$, $z_1 < z_2 < \cdots < z_n$ and n sequences $\langle z_j^i \rangle_{i=1}^{\infty}$, $j = 1, 2, \ldots, n$ such that

$$(1) \qquad z_j^i \to z_j, \quad 1 \le j \le n$$

$$(2) \qquad f(z_j^i) \ge p(z_j) - \frac{1}{i}, \quad j \text{ odd}, \ 1 \le j \le n, \ 1 \le i < \infty.$$

$$(3) \qquad g(z_j^i) \le p(z_j) + \frac{1}{i}, \quad j \text{ even}, \ 2 \le j \le n, \ 1 \le i < \infty.$$

The interpolation property of T guarantees that M is not empty. Define the functional F for each p in M by

$$F(p) = \max\{\sup\{p(x) - g(x) : x \in [a, b]\}, \sup\{f(x) - p(x) : x \in [a, b]\}\}$$

The theorem will be established if we can show that $F(p^*) = 0$ for some $p^* \in M$. Let $\rho = \inf\{F(p) : p \in M\}$ and choose $\langle p_k \rangle$ in M so that $F(p_k) \to \rho$. It can be shown that if $\langle \|p_k\| \rangle$ is unbounded, $\langle p_k(x) \rangle$ is unbounded at all but at most $n-1$ points in $[a, b]$, (see [5]), so we may assume $\langle p_k \rangle$ is bounded. Since T is finite dimensional, we may assume, by taking subsequences if necessary, that $p_k \to \bar{p} \in T$. We show that $\bar{p} \in M$. For each k, let $\langle z_{jk}^i \rangle_{i=1}^{\infty}$ and z_{jk}, for $j = 1, 2, \ldots, n$, be the n sequences and sequence limits associated with p_k from the definition of M. Assume without loss of generality that $|z_{jk}^i - z_{jk}| \le 1/i$ for all k and all j. Again by taking subsequences if necessary, we may assume that $z_{jk} \to \bar{z}_j$ for $1 \le j \le n$, where $\bar{z}_1 \le \bar{z}_2 \le \cdots \le \bar{z}_n$ are in $[a, b]$. Finally, we may assume without loss of generality that $|z_{jk} - \bar{z}_j| \le 1/k$ for all k and j. For each j, we find a sequence $\langle y_j^m \rangle_{m=1}^{\infty}$ and a sequence $\langle s_j^m \rangle_{m=1}^{\infty}$ such that $s_j^m \to 0$ and $y_j^m \to \bar{z}_j$ and either $f(y_j^m) \ge \bar{p}(\bar{z}_j) - s_j^m$ if j is odd or $g(y_j^m) \le \bar{p}(\bar{z}_j) + s_j^m$ if j is even. Indeed define $\langle y_j^m \rangle_{m=1}^{\infty}$ by $y_j^m = z_{jm}^m$ for all m. Then

$$|z_{jm}^m - \bar{z}_j| \le |z_{jm}^m - z_{jm}| + |z_{jm} - \bar{z}_j| \le \frac{1}{m} + \frac{1}{m} = \frac{2}{m}$$

so $y_j^m \to \bar{z}_j$ for each j. Suppose that j is odd. Then

$$f(z_{jm}^m) \geq p_m(z_{jm}) - \frac{1}{m} = \bar{p}(\bar{z}_j) - [(\bar{p}(\bar{z}_j) - \bar{p}(z_{jm})) + (\bar{p}(z_{jm}) - p_m(z_{jm}))] - \frac{1}{m}$$

Setting $s_j^m = [(\bar{p}(\bar{z}_j) - \bar{p}(z_{jm})) + (\bar{p}(z_{jm}) - p_m(z_{jm}))] + 1/m$, we observe that $s_j^m \to 0$ for each j by the uniform convergence of $\langle p_m \rangle$ to $\bar{p}$ and the continuity of $\bar{p}$. A similar treatment is used for j even. By taking subsequences if necessary, we may find sequences that satisfy the conditions in the definition of M. It is clear from the continuity of $\bar{p}$ that we in fact have $\bar{z}_1 < \bar{z}_2 < \cdots < \bar{z}_n$. Thus $\bar{p} \in M$. If $\rho = F(\bar{p}) = 0$ we are done, so suppose for a contradiction that $F(\bar{p}) > 0$. Using standard arguments and the fact that T is a Tchebycheff space, we may find n disjoint intervals $I_j = [a_j, b_j]$, $j = 1, 2, \ldots, n$ and a $\delta > 0$ such that $\bar{z}_j \in I_j$ and $\bar{p}(x) \leq g(x) - \delta$ if $x \notin \cup\{I_j : j \text{ is even}\}$ or $\bar{p}(x) \geq f(x) + \delta$ for $x \notin \cup\{I_j : j \text{ is odd}\}$. Define the n-tuple $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ by

$$\sigma_j = \begin{cases} 1 & \text{if } j \text{ is even and } \bar{p}(x) > g(x) \text{ for some } x \in I_j \\ -1 & \text{if } j \text{ is odd and } \bar{p}(x) < f(x) \text{ for some } x \in I_j \\ 0 & \text{otherwise} \end{cases}$$

Let $d_j = \frac{1}{2}(b_j + a_{j+1})$ for $j = 1, 2, \ldots, n-1$. It is possible to find an element h of T such that h has a nodal zero at each d_j for which $|\sigma_j - \sigma_{j+1}| \neq 1$ and has no other zeros in [a, b]. ([8], theorem 6.5). By supposition, for some $j = j^*$, $\sigma_j^* \neq 0$. For sufficiently small $\varepsilon > 0$, $\hat{p} = \bar{p} - \varepsilon\sigma_j^* \operatorname{sgn}(h(\bar{z}_j^*))h$ is such that $0 < \sup\{\hat{p}(x) - g(x) : x \in I_j^*\} < \sup\{\bar{p}(x) - g(x) : x \in I_j^*\}$ if j^* is even or $0 < \sup\{f(x) - \hat{p}(x) : x \in I_j^*\} < \sup\{f(x) - \bar{p}(x) : x \in I_j^*\}$ if j^* is odd. From the definitions of h and $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, we may further deduce by working step by step from I_j^* through the intervals to the right and left of I_j^*, that for sufficiently small $\varepsilon > 0$, inequalities of the above form hold for all I_j with $\sigma_j \neq 0$, that $\hat{p} \in M$, that $\hat{p}(x) < g(x)$ if $x \notin \cup\{I_j : j \text{ even}\}$ and $\hat{p}(x) > f(x)$ if $x \notin \cup\{I_j : j \text{ is odd}\}$. These facts together imply that $F(\hat{p}) < \rho$, a contradiction and the theorem follows.

We comment that if f and g are assumed continuous, the proof is substantially simplified.

The technique used above can also be employed to prove an old result of Davis [1], though it is not clear if the many more recent generalizations of this result can be similarly achieved.

REFERENCES

1. C. Davis, *Extrema of a polynomial, advanced problem 4714*, Amer. Math. Monthly, Nov. 1957, 679–680.

2. B. Gopinath and R. P. Kurshan, *The oscillation theorem for Tchebycheff spaces of bounded functions, and a converse*, J. of Approx. Theory **21** (1977), 151–173.

3. S. Karlin, *Representation theorems for positive functions*, J. Math. Mech. **12** (1963), 599–618.

4. S. Karlin and W. J. Studden, Tchebycheff Systems: With Applications in Analysis and Statistics, Interscience, New York, 1966.

5. L. Keener, *Existence of best uniform approximations by reciprocals*, J. of Approx. Theory **24** (1978), 245–250.

6. M. G. Krein and A. A. Nudel'man, *The Markov moment problem and extremal problems*, Vol. 50, Translations of Mathematical Monographs, A.M.S., Providence, R.I., 1977.

7. A. Pinkus, *Applications of representation theorems to problems of Chebyshev approximation with constraints*, Stud. Spline Approx. Theory (1976), 83–111.

8. R. Zielke, *Discontinuous Čebyšev system*, Lecture Notes in Mathematics no. 707, Springer-Verlag, 1979.

DEPARTMENT OF MATHEMATICS
DALHOUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA

and

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OREGON

Canad. Math. Bull. Vol. **25** (2), 1982

ON THE NUMBER OF PARTITIONS OF $\{1, \ldots, n\}$ INTO TWO SETS OF EQUAL CARDINALITIES AND EQUAL SUMS

BY
HELMUT PRODINGER

ABSTRACT. Let $A(n)$ be the number of partitions of $\{1, \ldots, n\}$ into two sets A, B of cardinality $n/2$ such that $\sum_{k\in A} k = \sum_{k\in B} k$. Then there is the asymptotic result

$$A(n) \sim \frac{2^n}{n^2}\frac{4\sqrt{3}}{\pi} \quad \text{as} \quad n \to \infty, \qquad n \equiv 0 \pmod 4).$$

1. Introduction. Suppose that the best n tennis players play a master tournament in such a way that, as a first step, two sets of $n/2$ players and equal power play two sub-tournaments.

In mathematical language this reads: The set $\{1, \ldots, n\}$ is partitioned into two sets A, B of cardinality $n/2$ such that

$$\sum_{k\in A} k = \sum_{k\in B} k = \frac{n(n+1)}{4}. \tag{1}$$

In this paper the number $A(n)$ of such partitions is considered. Apparently $n \equiv 0 \pmod 4$ must hold. For instance, for $n=4$ there are two solutions $A=\{1,4\}$, $B=\{2,3\}$ and $A=\{2,3\}$, $B=\{1,4\}$, hence $A(4)=2$.

An asymptotic answer is

THEOREM.

$$A(n) = \left|\left\{(\varepsilon_1, \ldots, \varepsilon_n) \,\middle|\, \varepsilon_i \in \{-1, 1\}, \sum_{k=1}^{n} \varepsilon_k = 0, \sum_{k=1}^{n} \varepsilon_k k = 0\right\}\right| \sim \frac{2^n}{n^2}\frac{4\sqrt{3}}{\pi}.$$

The proof of this result is along the lines of [1], where it is shown that

$$B(n) = \left|\left\{(\varepsilon_{-n}, \ldots, \varepsilon_n) \,\middle|\, \varepsilon_i \in \{0, 1\}, \sum_{k=-n}^{n} \varepsilon_k k = 0\right\}\right| \sim \frac{2^{2n+1}}{n^{3/2}}\sqrt{\frac{3}{\pi}}.$$

However, the present situation is more complicated.

2. Proof of the Theorem. $A(n)$ is the constant term in the expansion of $\prod_{k=1}^{n}(uz^k + u^{-1}z^{-k})$. This yields with $u = e^{ix}$, $z = e^{iy}$

$$\prod_{k=1}^{n}(uz^k + u^{-1}z^{-k}) = 2^n \prod_{k=1}^{n} \cos(x+ky) =: 2^n f_n(x, y).$$

Received by the editors May 7, 1980 and, in revised form, February 2, 1981.
AMS Subject Classification (1980): 05A17.

Note that

$$f_n(\pi+x, y)=f_n(x, y) \quad \text{if} \quad n\equiv 0, 2 \pmod 4$$

and

$$f_n(x, \pi+y)=f_n(x, y) \quad \text{if} \quad n\equiv 0, 3 \pmod 4.$$

Since $f_n(x, y)$ is just a trigonometrical polynomial, its constant term (which is $2^{-n}A(n)$) is found by integrating. Hence

$$\begin{aligned} 4\pi^2 2^{-n}A(n) &= \int_{-\pi/2}^{3\pi/2}\int_{-\pi/2}^{3\pi/2} f_n(x, y)\,dx\,dy \\ &= \int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2} \{f_n(x, y)+f_n(\pi+x, y)+f_n(x, \pi+y) \\ &\quad +f_n(\pi+x, \pi+y)\}\,dx\,dy \\ &= \int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2} 4f_n(x, y)\,dx\,dy, \quad \text{only if} \quad n\equiv 0 \pmod 4. \end{aligned}$$

Hence the condition $n\equiv 0 \pmod 4$ is assumed to hold throughout the rest of this paper.

Now the integrand will be estimated for values of y not near to the origin.

$$\begin{aligned} \left[\prod_{k=1}^{n}\cos(x+ky)\right]^2 &= \prod_{k=1}^{n}(1-\sin^2(x+ky)) \\ &< \exp\left[-\sum_{k=1}^{n}\sin^2(x+ky)\right] \qquad (2) \\ &= \exp\left[-\frac{n}{2}+\frac{\cos((n+1)y+2x)\cdot \sin ny}{2\sin y}\right]=0(e^{-\beta n}), \end{aligned}$$

with $\beta>0$ for $\pi/2(n+1)\le|y|\le\pi/2$. Hence the integration with respect to y is only to be done in the interval $[-\pi/2(n+1), \pi/2(n+1)]$:

$$\pi^2 2^{-n}A(n)\sim\int_{-\pi/2(n+1)}^{\pi/2(n+1)}\int_{-\pi/2}^{\pi/2} f_n(x, y)\,dx\,dy.$$

Now assume that $|(n+1)y+2x|\ge\pi/2$, $|y|\le\pi/2(n+1)$, $|x|\le\pi/2$ holds. Since $f_n(-x, -y)=f_n(x, y)$, it is sufficient to discuss

$$(n+1)y+2x\ge\frac{\pi}{2}, \qquad y\le\frac{\pi}{2(n+1)}, \qquad x\le\frac{\pi}{2}.$$

In the estimation (2), the cosine is negative and thus the integrand is again $0(e^{-\gamma n})$, $\gamma>0$. Hence

$$\pi^2 2^{-n}A(n)\sim\iint f_n(x, y)\,dx\,dy,$$

where the integration is to be done in the domain

$$D=\left\{(x,y)\,\middle|\,|x|\le\frac{\pi}{2},\ |y|\le\frac{\pi}{2(n+1)},\ |(n+1)y+2x|\le\frac{\pi}{2}\right\}.$$

Note that $-\pi/2\le x+ky\le\pi/2$ holds for $k=1,\ldots,n$ in this domain. Also note that $\cos z\le\exp(-\frac{1}{2}z^2)$ for $-\pi/2\le z\le\pi/2$. Thus

$$\begin{aligned}\pi^2 2^{-n}A(n)&\sim\iint_D f_n(x,y)\,dx\,dy\le\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\tfrac{1}{2}\sum_{k=1}^{n}(x+ky)^2\right]dx\,dy\\&\sim\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\tfrac{1}{2}\left(nx^2+n^2xy+\frac{n^2}{3}y^2\right)\right]dx\,dy\\&=\frac{1}{\sqrt{n}}\frac{\sqrt{3}}{n^{3/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-(x^2+2rxy+y^2)/2)\,dx\,dy\\&=\frac{\sqrt{3}}{n^2}\cdot 4\pi\qquad(\text{with } r=\sqrt{3}/2).\end{aligned}$$

Now for $|x|<n^{-1/3}$, $|y|<n^{-4/3}$:

$$\begin{aligned}\prod_{k=1}^{n}\cos(x+ky)&=\prod_{k=1}^{n}\exp(-\tfrac{1}{2}(x+ky)^2)\prod_{k=1}^{n}\{1+0((x+ky)^4)\}\\&=\exp\left[-\tfrac{1}{2}\sum_{k=1}^{n}(x+ky)^2+0(n^{-1/3})\right].\end{aligned}$$

Note that for $(x,y)\in D$ the integrand is positive. Hence

$$\begin{aligned}\iint_D f_n(x,y)\,dx\,dy&>\int_{-n^{-4/3}}^{n^{-4/3}}\int_{-n^{-1/3}}^{n^{-1/3}}f_n(x,y)\,dx\,dy\\&\sim\int_{-n^{-4/3}}^{n^{-4/3}}\int_{-n^{-1/3}}^{n^{-1/3}}\exp\left[-\tfrac{1}{2}\sum_{k=1}^{n}(x+ky)^2\right]dx\,dy\sim\frac{\sqrt{3}}{n^2}4\pi.\end{aligned}$$

Therefore $(\sqrt{3}\cdot 4\pi)/n^2$ is an asymptotic equivalent for $\pi^2 2^{-n}A(n)$.

3. Miscellaneous. Here are some numerical values.

n	$A(n)$	$\frac{2^n 4\sqrt{3}}{n^2\pi}$	$A(n)\Big/\frac{2^n 4\sqrt{3}}{n^2\pi}$
4	2	2.205316	0.9069
8	8	8.821264	0.9069
12	58	62.72899	0.9246
16	526	564.5609	0.9317
20	5448	5781.104	0.9424
24	61108	64234.48	0.9513
28	723354	755082.9	0.9580

In the language of probability theory, the theorem can be reformulated. If ε_k are independent identically distributed random variables with values -1 and 1, each with probability $\frac{1}{2}$, then

$$P\left(\sum_{k=1}^{n} \varepsilon_k k = 0 \quad\text{and}\quad \sum_{k=1}^{n} k = 0\right) \sim \frac{1}{n^2}\frac{4\sqrt{3}}{\pi}. \tag{3}$$

Now

$$P\left(\sum_{k=1}^{n} \varepsilon_k = 0\right) = \binom{n}{n/2} 2^{-n} \sim \sqrt{\frac{2}{\pi n}}, \tag{4}$$

by Stirling's formula. Furthermore,

$$P\left(\sum_{k=1}^{n} \varepsilon_k k = 0\right) \sim \frac{1}{n^{3/2}}\sqrt{\frac{6}{\pi}}, \tag{5}$$

which can be derived by van Lint's method [1].

It is worth noting that $f_n(x, y)$ is the characteristic function of the random vector $S = X_1 + \cdots + X_n$ where $X_k = \pm(1, k)$ each with probability $\frac{1}{2}$. From the Liapounov Central Limit Theorem it follows that if $S = (S_1, S_2)$ then $(S_1 n^{-1/2}, \sqrt{3} S_2 n^{-3/2})$ is asymptotically normally distributed with the density function

$$\frac{1}{2\pi\sqrt{1-r^2}} \exp\{-(t^2 - 2rtu + u^2)/2(1-r^2)\},$$

where $r = \sqrt{3}/2$ is the asymptotic correlation (see [2]).

Thus (3) is not unexpected; (3) is two times the product of (4) and (5) and this factor 2 is just $(1-r^2)^{-1/2}$.

I wish to thank F. J. Urbanek, F. Vogl and the referee for helpful remarks.

REFERENCES

1. J. H. van Lint, Representations of 0 as $\sum_{k=-N}^{N} \varepsilon_k k$, Proc. A.M.S. 18 (1967), 182–184.
2. J. V. Uspensky, "Introduction to Mathematical Probability", McGraw-Hill, New York (1937).

INSTITUT FÜR MATHEMATISCHE LOGIK
UND FORMALE SPRACHEN, TU WIEN
1040 WIEN GUßHAUSSTRAßE 27–29 AUSTRIA

Canad. Math. Bull. Vol. **25** (2), 1982

DISCRETE SETS AND DISCRETE MAPS

BY
RICHARD H. WARREN

ABSTRACT. A subset of a topological space is called discrete iff every point in the space has a neighborhood which meets the set in at most one point. Discrete sets are useful for decomposing the images of certain maps and for generalizing closed maps. All discrete sets are closed iff the space is T_1. As a result of characterizing discrete and countably discrete maps, theorems due to Vaĭnšteĭn and Engelking are extended to these maps.

In [A] it is proved that certain image spaces Y from a closed map f can be decomposed $Y = \bigcup_{n=0}^{\infty} Y_n$ where Y_n is discrete in Y for all $n \geq 1$ and $f^{-1}(y)$ is compact for all y in Y_0. Tani [T] showed that the same decomposition holds for discrete maps which are a generalization of closed maps. We extend Engelking's technique [E] for characterizing closed maps to obtain characterizations of discrete and countably discrete maps. Also, we prove that the set of points for which a map is discrete or countably discrete is a G_δ*-set.*

An arbitrary (countable) set in a topological space is called *discrete* (*countably discrete*) iff every point in the space has a neighborhood which meets the set in at most one point.

Since $\{x\}$ is discrete, it follows that a topological space is T_1 iff each discrete set has no limit points. Likewise, a space is T_1 iff each of its discrete sets is closed. Since each subset of a discrete set is discrete, in a T_1-space a set is discrete iff all of its subsets are closed. Thus, in a T_1-space a set is discrete iff it is closed and the relative topology on it is the discrete topology. Furthermore, a set is closed and discrete iff all of its subsets are closed. Therefore, a set is closed and discrete iff all of the supersets of its complement are open.

Since this last concept will be used to characterize discrete maps and determine when discrete maps are closed, we call a set in a topological space *superopen* iff all of its supersets are open; i.e., if W is a set in a space X, then W is superopen iff whenever $W \subset V \subset X$, it follows that V is open in the space. A set is called ω*-superopen* iff it is superopen and its complement is countable.

A function $f: X \rightarrow Y$ is called a *discrete* (*countably discrete*) map iff for every discrete (countably discrete) closed set A in X, $f(A)$ is closed in Y. We do not assume that a discrete map is continuous or onto, as Tani has. It is easily

---

Received by the editors, June 11, 1980 and in revised form, October 25, 1980.
AMS Subj. Class.: Primary 54C10, 54A25.
Key Words: discrete set, countably discrete set, generalizations of closed map.

verified that if X is a T_1-space and if $f: X \to Y$ is a countably discrete map onto Y, then Y is a T_1-space. If X is a T_1-space, $f: X \to Y$ is a discrete map and A is a discrete (countably discrete) subset of X, then $f(A)$ is a discrete (countably discrete) subset of Y.

Let $f: X \to Y$ be a function. Denote by $C(f)$ ($D(f), E(f)$) the subset of Y consisting of all points y such that for every open (superopen, ω-superopen) $W \subset X$ containing $f^{-1}(y)$, there is a neighborhood $N_y \subset Y$ satisfying $f^{-1}(N_y) \subset W$. Engelking [E] showed that f is closed iff $C(f) = Y$. Clearly $C(f) \subset D(f) \subset E(f)$. We note that if f maps a space X into a space Y, then f is a discrete (countably discrete) map iff $D(f) = Y$ ($E(f) = Y$). Also, if f is a continuous countably discrete map of a T_1-space into a sequential T_2-space, then f is a closed map and $C(f) = D(f) = E(f)$. Since a first countable space is a sequential space, it follows that we are able to state two theorems of Vaĭnšteĭn (see [E]) in terms of countable discrete maps.

Theorem 1. *For every continuous mapping $f: X \to Y$ of a complete metric space X into a first countable T_2-space Y and for any set $A \subset X$ such that $f \mid A : A \to f(A)$ is countably discrete, there exists a G_δ-set $B \subset X$ such that $A \subset B$ and $f \mid B : B \to f(B)$ is closed.*

Theorem 2. *The image of a complete metric space under a continuous countably discrete map has a complete metric if it is a first countable T_2-space.*

For the remainder of this paper we assume that $f: X \to Y$ is a continuous mapping of a metric space X with metric ρ onto a first countably T_2-space Y.

Theorem 3. *For every $y \in E(f)$, the boundary of $f^{-1}(y)$ is compact.*

Proof. Denote the boundary of $f^{-1}(y)$ by F. Let $\{V_n : n = 1, 2, \ldots\}$ be a base at y and let $A = \{x_1, x_2, \ldots\}$ be a countably infinite subset of F. Since $x_n \in F$, it follows that $x_n \in \overline{X - f^{-1}(y)}$, but $x_n \notin X - f^{-1}(y)$ as $f^{-1}(y)$ is closed and thus contains F. So for every n we can choose a point $z_n \in f^{-1}(V_n) - f^{-1}(y)$ such that $\rho(x_n, z_n) < 1/n$. Set $Z = \{z_1, z_2, \ldots\}$. As $y \in E(f)$, the set $X - Z \supset f^{-1}(y)$ is not superopen. Hence there is a set $B \subset Z$ which is not closed. Therefore $\emptyset \neq B' \subset Z'$, and so $A' \neq \emptyset$. Thus F is compact.

Corollary. *For every $y \in D(f)$, the boundary of $f^{-1}(y)$ is compact.*

Let $W_i(f)$ denote for $i = 1, 2, \ldots$ the subset of Y consisting of all points $y \in Y$ which have a neighborhood $N_y \subset Y$ such that every set $K \subset f^{-1}(N_y)$ satisfying the conditions

$$(1) \qquad \rho(s, t) \geq 1/i \quad \text{and} \quad f(s) \neq f(t) \quad \text{for distinct } s, t \in K$$

is finite. Clearly, the sets $W_i(f)$ are open.

Lemma 1. *$E(f) \subset W_i(f)$ for $i = 1, 2, \ldots$.*

Proof. If $a \in X$, then we designate $B(a, r) = \{x \in X : \rho(a, x) < r\}$. Suppose that there exists a point $y \in E(f) - W_i(f)$. Let $\{V_n\}_{n=1}^{\infty}$ be a base at y. Choose for $n = 1, 2, \ldots$ an infinite subset K_n of $f^{-1}(V_n)$ satisfying conditions (1). By the second part of (1), there exists an infinite set $L_n \subset K_n$ such that $f(x) \neq y$ for all $x \in L_n$ for $n = 1, 2, \ldots$. We will select a point $x_n \in L_n$ for $n = 1, 2, \ldots$ such that $A = \{x_1, x_2, \ldots\}$ is discrete. Choose $x_1 \in L_1$. Having chosen $x_k \in L_k$ satisfying $\rho(x_j, x_k) \geq 1/2i$ for $1 \leq j < k$, then every open ball $B(x_j, 1/2i)$ for $1 \leq j \leq k$ contains at most one member of L_{k+1}. Therefore we can choose $x_{k+1} \in L_{k+1}$ such that $\rho(x_j, x_{k+1}) \geq 1/2i$ for $1 \leq j \leq k$. Thus for every $x \in X$, the open ball $B(x, 1/4i)$ contains at most one member of A and so A is discrete. It follows that $W = X\text{-}A$ is ω-superopen and contains $f^{-1}(y)$. As $y \in E(f)$, we have $f^{-1}(V_n) \subset W$ for some n. But this is impossible, because $x_n \in A \cap f^{-1}(V_n)$.

Engelking [E, Lemma 3] proved that:

LEMMA 2. *If the metric* ρ *is complete, then* $\bigcap_{i=1}^{\infty} W_i(f) \subset C(f)$.

THEOREM 4. *If f is a continuous mapping of a complete metric space X onto a first countable* T_2*-space Y, then* $C(f) = D(f) = E(f)$ *and is a* G_δ*-set in Y.*

Proof. This result is a consequence of Lemmas 1 and 2.

REFERENCES

[A] A. Arhangel'skiĭ. *On closed mappings, bicompact spaces, and a problem of P. Aleksandrov*, Pacific J. Math. **18** (1966), 201–208.

[E] R. Engelking, *Closed mappings on complete metric spaces*, Fund. Math. **70** (1971), 103–107.

[T] T. Tani, *Some generalizations of closed maps, their properties and M-spaces*, Math. Japon. **20** (1975), 237–252.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF NEBRASKA AT OMAHA
OMAHA, NE 68182

Canad. Math. Bull. Vol. **25** (2), 1982

THE EXTENDED CENTER OF COPRODUCTS

BY

WALLACE S. MARTINDALE, 3rd

Abstract. The sole purpose of this paper is to prove that if R_1 and R_2 are algebras with 1 over a common field F, with each $(R_i : F) > 1$ and at least one $(R_i : F) > 2$, then the extended center of the coproduct $R_1 \coprod R_2$ is equal to F.

Let R_1 and R_2 be arbitrary algebras with 1 over a common field F, with each $(R_i : F) > 1$, and let $R = R_1 \coprod R_2$ denote the coproduct of R_1 and R_2 over F. We fix an F-basis $\{x_i\} \cup 1$ for R_1 and an F-basis $\{y_j\} \cup 1$ for R_2. The various finite products of x_i's and y_j's, in which the x_i's and y_j's alternate, together with 1, will be called monomials; the degree of a monomial is the total number of x_i's and y_j's in it (thus deg $1 = 0$). The set $\{U_k\}$ of all monomials forms an F-basis for R. For $f = \sum \beta_k U_k \in R$, $\beta_k \in F$, the support of f consists of all U_k's such that $\beta_k \neq 0$. The degree of f is the highest degree of any monomial in its support. Every $f \neq 0$ may be written uniquely as the sum of its homogeneous components $f = f_n + f_{n-1} + \cdots + f_0$, $f_n \neq 0$, where all monomials in the support of f_i have degree i. We shall generally denote the highest degree component of f by $\hat{f}$ (rather than by f_n). For $0 \neq n$ even we use the suggestive notation $\hat{f} = f_{xy} + f_{yx}$, where the monomials in the support of f_{xy} (resp. f_{yx}) begin with an x_i, (resp. y_j) and end with a y_j (resp. x_i). For n odd we write $\hat{f} = f_x + f_y$, where the monomials in the support of f_x (resp. f_y) begin and end with an x_i (resp. y_j). An element $f \neq 0$ of degree $n > 0$ is called pure if for n even $\hat{f} = f_{xy}$ or $\hat{f} = fgy_x$ and for n odd $\hat{f} = f_x$ or $\hat{f} = f_y$. A pair (f, g) is called pure interacting if f and g are pure and the monomials of $\hat{f}$ end in the same letter in which the monomials of $\hat{g}$ begin. It is clear that for $f, g \neq 0$, $\deg(fg) = \deg f + \deg g$ unless (f, g) is a pure interacting pair. It follows in particular that R is always a prime ring (if $fg = 0$ for nonzero f and g then $fzg \neq 0$ for z approximately chosen as x_1 or y_1). For prime rings R in general the notion of the extended center C was introduced in [3]. We refer the reader to that source for the basic construction and properties, emphasizing the key property that for $\lambda \in C$ there exists a nonzero ideal I_λ of R such that $\lambda a \in R$ for all $a \in I_\lambda$. It is well known that C is a field containing the ordinary center of R.

Received by the editors June 16, 1980 and in revised form, September 2, 1980.
AMS Subject Classifications (1980) 16A12, 16A08.
Key words and Phrases: coproduct; extended center.

It is our purpose in the present note to show that if at least one of $(R_i:F)>2$ then the extended center C of R is F itself. Aside from the feeling that it should be potentially useful to know precisely what the extended center is for certain important classes of prime rings, our result here is specifically needed in a joint paper with Susan Montgomery [4], in which we determine the so-called normal closure of coproducts of domains. First, however, we mention that the determination of C for the case where $(R_1:F)=(R_2:F)=2$ follows immediately from [1], p. 26, example 12.2. Here one writes $R_1=F[x]$, $x^2=\alpha x+\gamma$, $R_2=F[y]$, $y^2=\beta y+\delta$, and shows that the center of R is the polynomial ring $F[t]$, where $t=xy+yx-\alpha y-\beta x$, and that $1, x, y, xy$ form an $F[t]$-basis for R. R is thus a prime *PI*-ring and so by the sharper version of Posner's Theorem the extended center is simply the field of fractions $F(t)$ of $F[t]$. We also mention in passing that in this case R is definitely not a primitive ring, whereas in case one of $(R_i:F)>2$ we have Lichtmann's very pretty result [2] that R is always primitive.

Henceforward we assume at least one of $(R_i:F)>2$, which assumption is needed to establish

Lemma 1. *Let g and h be nonzero elements of R such that*

(a) *$gfh=hfg$ for all $f\in R$.*

Then $\deg g=\deg h$.

Proof. We set $n=\deg g$, $m=\deg h$, and first show that without loss of generality we may assume that n is odd and $m>0$. Indeed, we observe that property (a) is preserved on replacing g, h by ug, uh for any u. Thus if n is odd but $m=0$ we use $u=x_1y_1+y_1x_1$ whence by assumption $n+2=\deg(ug)=\deg(uh)=m+2$. In case n is even we use $u=x_1+y_1$, whence by assumption $n+1=\deg(ug)=\deg(uh)=m+1$. We next show that m must be odd. Indeed, if $g_x\neq 0$ we take $f=y_j$. Then the $(\ )_{xy}$ and $(\ )_{yx}$ terms of (a) respectively give $g_xy_jh_{xy}=0$ and $0=h_{yx}y_jg_x$; and so $h_{xy}=h_{yx}=0$, a contradiction to m even. Therefore we may assume both n and m are odd and, for sake of argument, $n\geq m+2$, with $g_x\neq 0$. If $h_x\neq 0$ we replace f by $y_jx_iy_j$ in (a) to obtain $g_xy_jx_iy_jh_x=h_xy_jx_iy_jg_x$. It follows that every monomial in the support of g_x has y_j in the $m+1$st position and x_i in the $m+2$nd position. But as $(R_1:F)>2$ or $(R_2:F)>2$ we can get the same result with $x_{i'}$ in place of x_i $(i\neq i')$ or $y_{j'}$ in place of y_j $(j\neq j')$, a contradiction. If $h_x=0$ then $h_y\neq 0$ which results in the contradiction $g_xh_y=0$.

Lemma 2. *Let g and h be pure and homogeneous of degree n, let s and t be pure and homogeneous of degree m, and suppose (g,s) and (h,t) are both noninteracting pairs such that $gs=ht$. Then for some $\alpha\in F$ $h=\alpha g$ and $s=\alpha t$.*

Proof. Writing $g=\sum\alpha_iU_i$, $s=\sum\beta_jV_j$, $h=\sum\gamma_kW_k$, $t=\sum\delta_lZ_l$ we have the equation

$$\sum_{i,j} \alpha_i\beta_j U_i V_j = \sum_{k,l} \gamma_k \delta_l W_k Z_l,$$

in which $\{U_i V_j\}$ are distinct monomials of degree $n+m$ and likewise $\{W_k Z_l\}$ are distinct monomials of degree $n+m$. This forces, via a suitable reordering, $U_i = W_i$ and $V_j = Z_j$, and accordingly we may write $h = \sum \gamma_i U_i$ and $t = \sum \delta_j V_j$. Therefore from $\sum \alpha_i \beta_j U_i V_j = \sum \gamma_i \delta_j U_i V_j$ we have $\alpha_i \beta_j = \gamma_i \delta_j$ for all i, j. In other words

$$\alpha = \frac{\gamma_i}{\alpha_i} = \frac{\beta_j}{\delta_j}$$

for all i, j, whence α is a constant, with $h = \alpha g$ and $s = \alpha t$.

We are now in a position to fulfill the purpose of this note.

THEOREM. *Let R_1 and R_2 be algebras with* 1 *over a common field F, with each $(R_i : F) > 1$ and at least one $(R_i : F) > 2$. Then the extended center C of $R = R_1 \coprod R_2$ is equal to F.*

Proof. We remark that the desired conclusion $C = F$ is equivalent to the assertion that any pair of nonzero elements g, h satisfying property (a), namely,

(a) $gfh = hfg$ for all $f \in R$,

must be F-dependent. Let g, h be such a pair. By Lemma 1 we have $n = \deg g = \deg h$. We claim that without loss of generality n is odd. Indeed, if n is even we set $u = x_1 + y_1$ and note that ug, uh are of degree $n+1$ and also satisfy (a). By assumption we then have $uh = \alpha ug$ for some $\alpha \in F$, whence $u(h - \alpha g) = 0$ from which one concludes $h = \alpha g$. Our claim is thereby established and so we may assume now that n is odd. We next aim to show that there is $\alpha \in F$ such that $\hat{h} = \alpha \hat{g}$. Indeed, we may assume $g_x \neq 0$ which forces $h_x \neq 0$ (since $h_x = 0$, with $f = 1$ in (a), would force the contradiction $g_x h_y = 0$). We set $f = y_1$ in (a) to obtain $(g_x y_1) h_x = (h_x y_1) g_x$. Now by Lemma 2 there is $\alpha \in F$ such that $h_x = \alpha g_x$. If $g_y = 0$ (and hence $h_y = 0$) we are done. If $g_y \neq 0$ then the preceding argument shows there is $\beta \in F$ such that $h_y = \beta g_y$. On the other hand, setting $f = 1$ in (a) we get $g_x h_y = h_x g_y$, whence $g_x(\beta g_y) = (\alpha g_x) g_y$. It follows that $\alpha = \beta$ and so $\hat{h} = \alpha \hat{g}$. To complete the proof of the theorem we note $h - \alpha g$, if not already zero, has smaller degree than g. But it is also clear that the pair g, $h - \alpha g$ again satisfies (a), thus giving a contradiction to Lemma 1.

COROLLARY. *Let $\{R_\alpha \mid \alpha \in A, (R_\alpha : F) > 1, |A| > 1\}$ be a collection of algebras with* 1 *over a common field F, with either some $(R_\alpha : F) > 2$ or $|A| > 2$, and let $R = \coprod_\alpha R_\alpha$. Then the extended center of R is equal to F.*

ACKNOWLEDGEMENT. We are indebted to the referee for several suggestions toward simplifying and clarifying portions of this paper.

REFERENCES

1. G. Bergman, *Modules over coproducts of rings*, Trans. Amer. Math. Soc. **200** (1974), 1–32.

2. A. I. Lichtman, *The primitivity of free products of associative algebras*, J. Algebra **54** (1978), 153–158.

3. W. S. Martindale, 3rd, *Prime rings satisfying a generalized polynomial identity*, J. Algebra **12** (1969), 576–584.

4. W. S. Martindale, 3rd and M. S. Montgomery, *The normal closure of coproducts of domains*, submitted for publication.

UNIVERSITY OF MASSACHUSETTS,
AMHERST, MA 01003

Canad. Math. Bull. Vol. **25** (2), 1982

PROBLEMS AND SOLUTIONS

This department welcomes problems believed to be new. Normally, they should involve techniques not beyond the masters' level, not be too specialized and should require some ingenuity. However, attractive unsolved problems will be considered. If any problem comes from a lemma used in research or if it arises in an application, please include some remarks on the context.

Occasionally, we would like to feature a Canadian mathematics department. Accordingly, problem buffs at any place are invited to produce a joint slate of problems for one issue of the Bulletin. It is hoped that teachers may be able to glean from the Bulletin nice problems for graduate or advanced undergraduate classes.

Send all communications concerning this department to

PROBLEMES ET SOLUTIONS

Cette rubrique présente des problèmes inédits assez ingénieux mais pas trop complexes, du niveau du 2*e* cycle. La rédaction recevra aussi les problèmes intéressants non résolus. Les correspondants replaceront dans leurs contextes les problèmes découlant de nouveaux résultats de recherche ou survenant dans les applications.

Nous aimerions publier de temps à autre une chronique sur les mathématiques canadiennes. Nous invitons donc les fervents de problèmes de chacun des départments de mathématiques au Canada à nous fournir pour une même livraison une collection de problèmes de leur département. Enfin nous souhaitons que les professeurs trouveront dans cette chronique des problèmes pour leurs étudiants diplômés et pré-diplômés chevronnés.

Prière d'adresser problèmes et solutions à

E. J. Barbeau, Problems Editor
Canadian Mathematical Bulletin
Department of Mathematics
University of Toronto
Toronto, Ontario M5S 1A1

NEW PROBLEMS

P.320. Put $f_n = 2^n + 1$ for each positive integer n. Prove that f_5 and f_6 are the only two consecutive terms of the sequence $\{f_n\}$ which are products of two distinct primes that differ by the *same* amount (for this particular pair, the difference is 8).

Albert A. Mullin
Huntsville, Alabama

P.321. Put $g_n = 2^{2n-1} - 1$ for each positive integer n. Prove or disprove: g_5 and g_6 are the only two consecutive terms of the sequence $\{g_n\}$ which are products of two distinct primes that differ by the *same* amount (66 for this particular pair).

Albert A. Mullin
Huntsville, Alabama

P.322. Prove or disprove:

$$\lim_{n\to\infty}\left(\sum_{k=1}^{n}(-1)^{k-1}\frac{k}{2^k-1}\binom{n}{k}\right)=\frac{1}{\log 2}.$$

George Szekeres
University of New South Wales
Kensington, Australia

P.323. Let p and q be distinct primes such that $p\equiv q\equiv 5 \pmod 8$ and $(p/q)=-1$ (Legendre symbol). By the law of quadratic reciprocity we have $(2p/q)=(2q/p)=+1$, so that the biquadratic symbols $(2p/q)_4$ and $(2q/p)_4$ are well-defined: that is

$$(2p/q)_4=\begin{cases}+1 & \text{if } 2p \text{ is a biquadratic residue} \pmod q\\ -1 & \text{otherwise,}\end{cases}$$

with a similar meaning for $(2q/p)_4$. Determine the value of the product $(2p/q)_4(2q/p)_4$ in terms of the representations $p=a^2+b^2$, $q=c^2+d^2$. (This extends Burde's rational biquadratic reciprocity law; see K. Burde, *Ein rationales biquadratisches Reziprozitätsgesetz*, J. Reine und Angewandte Math. 235 (1969), 175–184.)

Kenneth S. Williams
Carleton University, Ottawa

P.324. Find the set of all real numbers p for which the following holds: suppose that $\{c_n\}$ is a positive sequence for which $\{n^{-1}\sum_{k=1}^{n}c_k\}$ is convergent; then

$$\left\{\left(\sum_{k=1}^{n}k^p\right)^{-1}\left(\sum_{k=1}^{n}c_kk^p\right)\right\}$$

is also convergent with the same limit.

F. S. Cater
Portland State University
Portland, Oregon

PROBLEMS IN NUMBER THEORY

Comments on these problems, posed at the 1981 Summer Meeting of the C.M.S. might be sent to the problem editor or to Richard K. Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta T2N 1N4.

R.7. Do there exist distinct primes $p_1, p_2, \ldots, p_r, q_1, q_2, \ldots, q_s$ such that $(\sum_{i=1}^{r} p_i^{-1})(\sum_{j=1}^{s} q_j^{-1}) = 1$? In particular, do there exist distinct primes $p, q_1, \ldots, q_s$ with $(\frac{1}{2} + (1/p))(\sum_{j=1}^{s} q_j^{-1}) = 1$?

E. J. BARBEAU

R.8. Consider the system of n congruences $x \equiv -4x_i \pmod{x_i^2}$ where $x_1, x_2, \ldots, x_n$ are pairwise coprime integers. By the Chinese Remainder Theorem, there is a solution between 1 and $x_1^2 x_2^2 \cdots x_n^2$ and this has the form $x = tx_1 x_2 \cdots x_n$. When does it happen that $t = 1$?

E. J. BARBEAU

R.9. Let $\mu(n)$ be the Möbius function. Are there infinitely many n with $\mu(n) = \mu(n+1) = 1$ or with $\mu(n) = \mu(n+1) = -1$.

M. V. SUBBARAO

SOLUTIONS

P.286. (a) Must every finite group of even order exceeding 2 have an automorphism of order 2?

(b) Must every finite group of odd order have an automorphism of order 2?

(c) Let p be a prime. Must every p-group of order exceeding 2 have an automorphism of order 2?

OTTO GEORGE
TADDLE CREEK, ONTARIO

Comment from Otto H. Kegel, Mathematisches Institut, Albert-Ludwigs-Universität, Freiburg. H. Heineken and H. Liebeck (Arch. der Math. 25 (1974), 8–16) give examples of finite p-groups, p odd, the automorphism group of which is a p-group.

P.299. Observe that $720 = 27^2 - 3^2 = 28^2 - 8^2 = 29^2 - 11^2$. Show that there are infinitely many positive integers which differ from the next three perfect squares greater than it by a perfect square. Are there infinitely many positive integers which differ from each of the next four greater perfect squares by a perfect square?

OTTO GEORGE
TADDLE CREEK, ONTARIO

Solution by the proposer. The answer to the question is yes. Any number n which differs from the next four squares by squares must satisfy, for suitable integers x, a, b, c: $(x-1)^2 \leq n = x^2 - a^2 = (x+1)^2 - b^2 = (x+2)^2 - c^2 = (x+3)^2 - d^2$. This implies that $b^2 - a^2$, $c^2 - b^2$, $d^2 - c^2$ are consecutive odd numbers and that $2x - 1 \geq a^2$. It suffices to determine four non-consecutive numbers a, b, c, d for which (1) $2(b^2+1) = a^2 + c^2$, (2) $2(c^2+1) = b^2 + d^2$, (3) $b^2 \geq 2(a^2+1)$, and determine x and n from these. Solutions to (1) and (3) are legion. One way to find a set is to begin with a number for which -1 is a quadratic residue along with the square root of -1 modulo that number; by way of example, pick 5 and 2 and set $b = 5k+2$. Then $b^2 + 1 = 5(5k^2 + 4k + 1)$ whence $2(b^2+1) = (k-1)^2 + (7k+3)^2$. The triple $(a, b, c) = (k-1,\ 5k+2,\ 7k+3)$ satisfies (1) and (3) for every positive integer k. To satisfy (2) as well, we require $d^2 = 73k^2 + 64k + 16$, which is equivalent to $(*)\ -73d^2 + z^2 = -144$, where $\pm z = 73k + 32$. Equation $(*)$ has infinitely many solutions (z_h, d_h) where (z_1, d_1) is either $(32, 4)$ or $(41, 5)$ and

$$z_{h+1} + \sqrt{73}d_{h+1} = (z_1 + \sqrt{73}d_1)(p_h + \sqrt{73}q_h)$$

with the solutions (p_h, q_h) of $p^2 - 73q^2 = 1$ given by $p_h + q_h\sqrt{73} = (2\,281\,249 + 267\,000\sqrt{73})^h$. In order to obtain integral k, we need values of z congruent to 32 or 41 modulo 73. Using $2\,281\,249 \equiv -1$ and $267\,000 \equiv 39 \pmod{73}$, it can be shown by induction that $p_{2j} \equiv 1$ and $q_{2j} \equiv -5j \pmod{73}$, whereupon $p_{146j} \equiv 1$ and $q_{146j} \equiv 0 \pmod{73}$ for $j = 1, 2, 3, \ldots$. Thus, any $z = z_h$ with $h \equiv 1 \pmod{146}$ will yield an integer k from which a solution of (1), (2), (3) is constructed.

P.300. Let $\{b_n\}$ be a sequence of real nonzero terms for which $\lim b_n = 0$. The function $f(x) = \prod_{n=0}^{\infty} \cos(b_n x)$ is well-defined for real values of x. Is it necessarily true that $\lim_{x\to\infty} f(x) = 0$?

PETER BORWEIN AND JON BORWEIN
DALHOUSIE UNIVERSITY, HALIFAX
CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PA

Solution by A. Meir, University of Alberta. The answer is no. Let $b_n = 2\pi/8^n$ $(n = 0, 1, \ldots)$ and $x_k = 8^k$ $(k = 1, 2, \ldots)$. then $f(x_k) = \prod_{j=1}^{\infty} \cos(2\pi/8^j) > \prod_{j=1}^{\infty} (1 - 2\pi^2/8^{2j}) = c > 0$. Thus $\limsup_{x\to\infty} f(x) > 0$.

Also solved by M. B. Gregory, University of North Dakota, Grand Forks; Ivo Klemes, student, University of Toronto; and the proposers. The proposers remark that it is a consequence of a theorem of Weyl (given by M. Kac in *Statistical independence in probability, analysis and number theory*, p. 45) that

$f(x)$ tends to zero in integral mean. Meir observes that $\lim f(x)=0$ is certainly possible; for example, let $b_n=n^{-1}$. He also poses the related question: is it possible for $\limsup f(x)$ to be equal to 1 as $x \to \infty$?

P.301. Let $\{a_n\}$ be any bounded sequence of nonnegative numbers. For each positive ε and each pair n, m of positive integers with $n<m$, define

$$E(\varepsilon)=\{j : a_j \geq \varepsilon\},$$

$$E(\varepsilon, n, m)=\{j : n \geq j<m \quad \text{and} \quad a_j \geq \varepsilon\},$$

and

$$dE(\varepsilon)=\liminf_{m\to\infty} m^{-1}\,|E(\varepsilon, 1, m)|.$$

($|S|$ denotes the number of elements in the set S.)

Prove that a necessary and sufficient condition that there exists a nonincreasing positive sequence $\{b_n\}$ for which $\sum b_n=\infty$ and $\sum a_n b_n<\infty$, is that $dE(\varepsilon)=0$ for all $\epsilon>0$.

F. S. Cater
Portland State University
Portland, Or

Solution. Suppose $dE(\varepsilon)=0$ for all $\varepsilon>0$. It is then possible to construct a sequence $\{m_k\}$ of integers for which $1=m_0<m_1<m_2<m_3<\cdots$, $m_j>2m_{j-1}$ and $|E(2^{-j}, m_{j-1}, m_j)|<2^{-j}(m_j-m_{j-1})$ for all $j>0$. Let $b_n=(m_j-m_{j-1})^{-1}$ for $m_{j-1}\leq n<m_j$. Then the sequence $\{b_n\}$ has the required properties.

On the other hand, suppose for some positive ε, it is true that $dE(\varepsilon)>0$. Let $E=E(\varepsilon)$ and suppose that $\{b_n\}$ is a nonincreasing positive sequence for which $\sum a_n b_n$ converges. Then $\sum\{b_n : n\in E\}\leq \varepsilon^{-1}\sum\{a_n b_n : n\in E\}<\infty$. We show that this implies the convergence of $\sum b_n$ and thus the necessity of the condition.

If E contains every integer some point on, there is nothing to prove. Otherwise, we can write the set of natural numbers as a disjoint union of finite consecutive intervals $E_1\cup F_1\cup E_2\cup F_2\cup E_3\cup F_3\cup\cdots$ with $E\cup\{1\}=E_1\cup E_2\cup E_3\cup E_4\cup\cdots$. Let $u_i=|E_i|$, $v_i=|F_i|$ and $x_i=\min\{b_j ; j\in E_i\}$, so that $\{x_i\}$ is nonincreasing, $\sum u_i x_i\leq\sum\{b_j : j\in E\cup\{1\}\}$ and $\sum v_i x_i\geq\sum\{b_j : j\in\bigcup F_i\}$. Since dE is positive, there is a number η with $0<\eta<1$ and a positive integer K for which $u_1+u_2+\cdots+u_k\geq\eta(u_1+v_1+\cdots+u_k+v_k)$ when $k\geq K$. Thus, when $k\geq K$, $(v_1+\cdots+v_k)\leq M(u_1+\cdots+u_k)$ where $M=\eta^{-1}(1-\eta)$.

Copyrighted material

Now, for each positive integer $p \geq K$,

$$\begin{aligned}
\sum_{i=1}^{p} v_i x_i &= (v_1 + \cdots + v_p)x_p + (v_1 + \cdots + v_{p-1})(x_{p-1} - x_p) \\
&\qquad + \cdots + (v_1 + \cdots + v_K)(x_K - x_{K+1}) + \sum_{j=1}^{K-1} (v_1 + \cdots + v_j)(x_j - x_{j+1}) \\
&\leq M(u_1 + \cdots + u_p)x_p + \sum_{j=1}^{p-i} M(u_1 + \cdots + u_j)(x_j - x_{j+1}) \\
&\qquad + \sum_{j=1}^{K-1} (v_1 + \cdots + v_j)(x_j - x_{j+1}) \\
&= M \sum_{i=1}^{p} u_i x_i + \sum_{j=1}^{K-1} (v_i + \cdots + v_j)(x_j - x_{j+1}),
\end{aligned}$$

and the required result follows.

Solved by the proposer, whose proof of sufficiency appears above.

P.302. Let X be the linear space $\mathbf{R}^2$ endowed with the norm $\|(x, y)\|_1 = |x| + |y|$, and let Y be the linear space $\mathbf{R}^2$ endowed with the norm $\|(x, y)\| = \max\{|x|, |y|\}$. Show that X and Y are isometric. Is this result true if $\mathbf{R}^2$ is replaced by $\mathbf{C}^2$?

George M. Rassias
University of Athens, Greece

Solution by the proposer. The mapping $(x, y) \rightarrow (x+y, x-y)$ is an isometry from $(\mathbf{R}^2, \|.\|_1)$ onto $(\mathbf{R}^2, \|.\|)$.

The set of extreme points of the unit ball in $(\mathbf{C}^2, \|.\|_1)$ is the union of two disjoint circles, namely $\{(z, 0) : |z| = 1\} \cup \{(0, w) : |w| = 1\}$. The set of extreme points of the unit ball in $(\mathbf{C}^2, \|.\|)$ is the torus $\{(z, w) : |z| = 1, |w| = 1\}$. Since the two norms define the same topology while the two sets of extreme points are not homeomorphic, $(\mathbf{C}^2, \|\cdot\|_1)$ and $(\mathbf{C}^2, \|\cdot\|)$ are not isometric.

P.303. Let R be a commutative ring with identity, J an ideal of R, and M an R-module. Let N be the submodule of M consisting of those elements x that are annihilated by some power (depending on x) of J.

(a) Prove that, if J is finitely generated, then J annihilates only the zero element of M/N.

(b) Find an example to show that the converse of part (a) is false.

(c) Find an example to show that the hypothesis "J is finitely generated" in part (a) cannot be dropped.

JAMES HEIN AND F. S. CATER
PORTLAND STATE UNIVERSITY
PORTLAND, OR

Solution by Alberto Facchini, Universita di Padova, Italy and A. A. Jagers, Technische Hogeschool Twente, Enschede, Netherlands (independently).

(a) Let the generators of J be $r_1, r_2, \ldots, r_t$. If $x+N$ is annihilated by J, then $r_ix \in N$, so that there exist $n_1, n_2, \ldots, n_t$ for which $J^{n_i}r_ix=0$ $(1 \le i \le t)$. If $n=\max(n_1, \ldots, n_t)$, then $J^{n+1}x=0$ so $x \in N$.

(b) Let R be any non-Noetherian integral domain and J an ideal of R which is not finitely generated. Take $M=R$. Then $N=\{0\}$ and J annihilates only the zero element of M/N since R has no zero divisors.

(c) Take $R=F[x_1, x_2, \ldots]$, the ring of all polynomials in countably many indeterminates over a field F, and let J be the maximal ideal generated by all the x_n. Then $I=x_1J+x_2J^2+\cdots$ is an ideal and $M=R/I$ is an R-module. It follows that $N=J/I$ and M/N is annihilated by J.

Also solved by the proposers.

Copyrighted material

RECOMMENDATIONS AUX AUTEURS

Pour permettre une publication plus rapide, tout article proposé pour publication doit se conformer aux indications suivantes:

1. Les articles doivent en principe ne pas dépasser 15 pages dactylographiées en double interlinge sur des feuilles de format 21.7 cm × 28 cm ($8\frac{1}{2}$" × 11"). Les articles plus longs peuvent être soumis au Journal Canadien de Mathématiques, l'Université Toronto, Toronto, Ontario, M5S 1A1.

2. Il faut numéroter toutes les pages. Donner sur la première page (i) un titre abrégé, (ii) un resumé ne depassant pas 100 mots et qu'on peut lire indépendamment de l'article, et (iii) la classification selon la "1980 Mathematics Subject Classification" (voir 1978 Subject Index, Mathematical Reviews) indiquant le sujet principal et les sujets sécondaires de l'article. L'affiliation institutionelle et l'addresse de chaque auteur doivent être placés au bas de la dernière page.

3. Toutes annotations, les caractères speciaux, les instructions particulières destinées à l'imprimerie et tous les caractères manuscrits doivent être catalogués et expliqués sur une feuille à part, ainsi que les codes de couleurs pour les annotations. Les caractères gothiques, les lettres grécques, et les caractères gras doivent être clairement indiqués avec soulignes en couleur. Toute ambiguité entre le chiffre 0 et la lettre o, miniscule ou majuscule, ainsi qu'entre le chiffre 1 et la lettre l doit être soigneusement evitée. Autant que possible on se conformera aux signes conventionnels en usage dans l'imprimerie. Les directives et les éxplications pour l'imprimeur doivent être portées sur une feuille séparée, attachée au manuscript.

4. Les figures seront tracées à l'encre de chine noire sur papier blanc glacé. Le nom de l'auteur et le titre de l'article devront être écrits au verso.

5. Les notes au bas des pages, autres que ceux qui rapportent au titre ou à l'auteur, doivent être enumerées de suite.

6. La bibliographie doit se trouver à la fin de l'article en ordre alphabétique. On se conformera aux abréviations conventionnelles (voir Mathematical Reviews, Annual Index) pour les titres des journaux. Exemples:

1. Poincaré, Henri, Sur un théorème de Geometrie, *Rend. Circ. Mat. Palermo* (1) 33 (1912) 375-407.

2. Titchmarsh, E.C., "The Theory of Functions", Oxford University Press, Oxford, 1932.

For an English version of the above see the issue of March or September.

Publications

available from
The Canadian Mathematical Society

Proceedings of the 13th Biennial Seminar 1971* Differential geometry and applications. Edited by J. R. Vanstone. 327 pages. ($14.00)

Proceedings of the 12th Biennial Seminar 1969* Time series, stochastic processes, convexity and combinatorics. Edited by R. Pyke. 318 pages. ($18.00)

Proceedings of the Fifth Canadian Mathematical Congress, Montreal 1961** Edited by E. M. Rosenthall. 230 pages. ($9.00)

Proceedings of the Fourth Canadian Mathematical Congress, Banff 1957** Edited by M. S. MacPhail. 192 pages. ($9.00)

Proceedings of the International Congress of Mathematicians, Vancouver 1974* 2 volumes. 1192 pages. ($67.50)

*Counting Labelled Trees** Lectures given at the 12th Biennial Seminar. By J. W. Moon. 123 pages. ($7.00)

*The Jeffery-Williams Lectures** (1968–1972) Lectures by I. Kaplansky, R. Pyke, W. A. J. Luxemburg, W. T. Tutte and P. J. Davis. 130 pages. ($7.00)

*Trigonometric Series*** A survey by R. L. Jeffery. 39 pages. ($4.00)

All prices are in Canadian dollars and include handling charges. Information on other C.M.S. publications can be obtained by writing the Executive Director at the address given below.

Please send order with remittance to:

Canadian Mathematical Society
Executive Office
577 King Edward
Ottawa, Ontario, Canada K1N 6N5

---

* A 20% discount is given to members of recognized mathematical societies.
** Limited availability.